FRAKE, APRIL N., M.A. Habitat Suitability and Ecological Niche Profiling of the West Nile Virus Vector, *Culex pipiens*, in Forsyth County, NC. (2014)
Directed by Dr. Roy Stine 75 pp.

Thought to have originated in Uganda in the late 1930's, West Nile Virus (WNV) was introduced in to North America in 1999 in New York City (Nash et. al, 2001). From its first occurrence within the United States, the virus spread across the contiguous forty-eight states and southern Canada in five years resulting in 18,000 human cases and over 700 fatalities (West Nile Virus, 2013). An important vector for the transmission of the WNV, *Culex pipiens* is widely distributed throughout the world with the exception of Australia and Antarctica (Farajollahi et al., 2011). This study seeks to utilize Remote Sensing, GIS, and Maximum Entropy (MaxEnt) Modeling in developing a presence-only habitat probability model of the known WNV bridge vector, *Cx. Pipiens* in Forsyth County, North Carolina by defining ecogeographical parameters that promote the mosquito species' larval development. Mosquito sampling was conducted in sixty-nine localities across the study area over a twenty-eight week period during the 2013 breeding season (April to October). Final habitat suitability maps produced as a result of this research will serve to guide future trap placement toward areas of high *Cx. pipiens* presence throughout the study area in an effort to optimize vector control measures and reduce the risk of WNV transmission. MaxEnt modeling results for the predicted probability of *Cx. pipien*s geographical distribution in Forsyth County highlighted the largest concentrations of *Cx. Pipiens* habitats within and along the periphery of the Winston-Salem municipality. Secondary areas of higher probability were located in the

north central portion of the county, an area marked by irrigated cropland and deciduous forest.

HABITAT SUITABILITY AND ECOLOGICAL NICHE PROFILING OF THE WEST NILE VIRUS VECTOR, *CULEX PIPIENS*, IN FORSYTH COUNTY, NC

by

April N. Frake

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2014

Approved by

Dr. Roy Stine
Committee Chair

APPROVAL PAGE

This thesis written by April N. Frake has been approved by the following committee of the Faculty of The Graduate School at the University of North Carolina at Greensboro.

Committee Chair	Dr. Roy Stine
Committee Members	Dr. Zhi-Jun Liu
	Dr. Ping Yin

_____April 2, 2014_____
Date of Acceptance by Committee

_____April 2, 2014_____
Date of Final Oral Examination

ACKNOWLEDGEMENTS

The success of this thesis is a result of the expertise, mentorship, and support of my committee who assisted with facilitating critical discussion, data processing, analysis, technical understanding, and writing support. First, to my adviser, Dr. Roy Stine who afforded me the freedom to explore new techniques, methodologies, and assistance in fusing my passions for Remote Sensing, GIS, and now, Medical Geography- I am deeply grateful for your encouragement and mentorship throughout this process. To Dr. Zhi-Jun Liu who instilled within me a passion for spatial statistics and worked alongside me through countless phases of planning and execution of the finer points of the methodologies explored herein- I am immensely grateful. Also, to Dr. Ping Yin who assisted me in developing a broader technical understanding of GIS and aided in formulating workflows that ultimately led this project to a successful completion- your efforts and support have been greatly appreciated.

Beyond my committee members, I am indebted to two other individuals who served as outside advisers to this project. First, to Dr. Mary Hall-Brown (Beebe), who leant her assistance in innumerable ways along this journey- I am richly blessed by your friendship and to have had the opportunity to work and learn from you. Also to Ryan Harrison with Forsyth County Vector Control who provided the collection data for this study in addition to introducing me to morphological identification of mosquito species and a wealth of first-hand knowledge on *Cx. pipiens* and a variety of other vectors present within the study area.

A special thank you is extended to Jim Nelson, whose technical support and assistance helped see this project to completion- I am immensely grateful for your willingness to provide me with the software upgrades, storage capacity, and assistance with countless technical requests and queries along the way. Thank you for always being incredibly accommodating.

To my fellow graduate student colleagues both in and outside the discipline of Geography who facilitated in critical discussion, assisted in data processing, and worked alongside me to create more refined workflows, in no particular order- Chet Gnegy, Jacob Turner, John Nowlin, Doug Gallaway and Megan Grigg- I am thankful to each one of you and look forward to being a part of your future successes.

Finally, to my Husband, Adam Frake whose unconditional love and unwavering support have sustained me throughout the duration of my academic career- I will never be able to fully express to you just how much the success I have experienced is as much at the hands of my effort as it is your support. I love you. None of this would have been possible without you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

West Nile Virus (WNV) is a vector-borne infectious disease that causes febrile illness, meningitis, encephalitis, and in less than one-percent of cases, death in humans (West Nile Virus, 2013). First discovered in 1937 from a native woman of the West Nile province of Uganda, WNV was introduced in to North America in 1999 in New York City (Dauphin et. al, 2004, Nash et. al, 2001). From its introduction into the United States in 1999 to 2012, a total of 37,088 cases were reported to the Centers for Disease Control; of these, 1,549 resulted in death.

The continued annual rise in WNV cases is cause for concern. In North Carolina, forty-three species of mosquitoes are known carriers of WNV; yet, the virus is predominately transmitted by the *Culex* mosquito including the *Culex pipiens* species (Andreadis et al., 2004; Hamer et al., 2008; Kilpatrick et al., 2005; Turell et al., 2002; West Nile Virus, 2013). In 2012, seven cases of WNV were reported in seven counties across the state of North Carolina (N.C. Department of Health and Human Services, 2013). As annual budget restraints continue to impact Vector Control divisions across the state, predictive mapping of *Cx. pipiens* populations could improve surveillance and prevention measures in areas of high vector abundance as mosquito-borne disease

transmission has been shown to be closely tied to behavior and population dynamics (Bolling et al., 2009, Ebel et al., 2005).

The objective of this research focuses on the creation of a predictive model for the presence of *Cx. pipiens* habitats utilizing Geographic Information Systems (GIS), Remote Sensing, and Maximum Entropy (MaxEnt) Modeling methodology in Forsyth County, North Carolina. Though an imperfect understanding of species distribution of *Cx. pipiens* still exists, species are expected to be non-randomly distributed across a variety of ecological settings directly related to their biological characteristics and tolerance toward deviations from optimal conditions (Hutchinson, 1957; Hirzel, 2002). As most zoonotic vectors are commonly associated with landscape and environmental determinants that directly impact their distribution and abundance (Brown et al., 1995), the modeling approach undertaken in this research will relate field observations of species occurrence to a set of ecogeographical variables used as predictors of environmental suitability. Final habitat suitability maps produced as a result of this research will serve to guide future trap placement toward areas of high *Cx. pipiens* presence throughout the study area in an effort to optimize Vector Control measures and reduce the risk of WNV transmission.

CHAPTER II

BACKGROUND

2.1 West Nile Virus

West Nile Virus (WNV) is a mosquito-borne flavivirus that is transmitted by the bite of infected female mosquitoes that acquire the virus from infected birds (Dauphin et al., 2004). As the virus is ingested during bloodmeal, it begins to invade the host's midgut cells and subsequently replicates and spreads to other tissue over the course of several days. Infectious mosquitoes carrying virus in their salivary glands will infect other vertebrate animals or mammals when injecting saliva and other chemically complex mixtures of anticoagulants and immune modulating factors during bloodmeal (Green & Reid, 2013).

Infected mosquitoes possess the ability to transmit the virus to a wide range of vertebrate hosts including humans, birds, horses, and other mammals (West Nile Virus, 2013). While the aforementioned hosts do not produce significant levels of viraemia (the concentration of virus in the blood) to contribute to the furthered transmission cycle of the virus, many species of birds serve as critical hosts in the WNV cycle (Figure 1) (Dauphin et al., 2004). In the United States, WNV has been detected in over 300 species of dead birds including crows, blue jays, and sparrows (West Nile Virus, 2013).

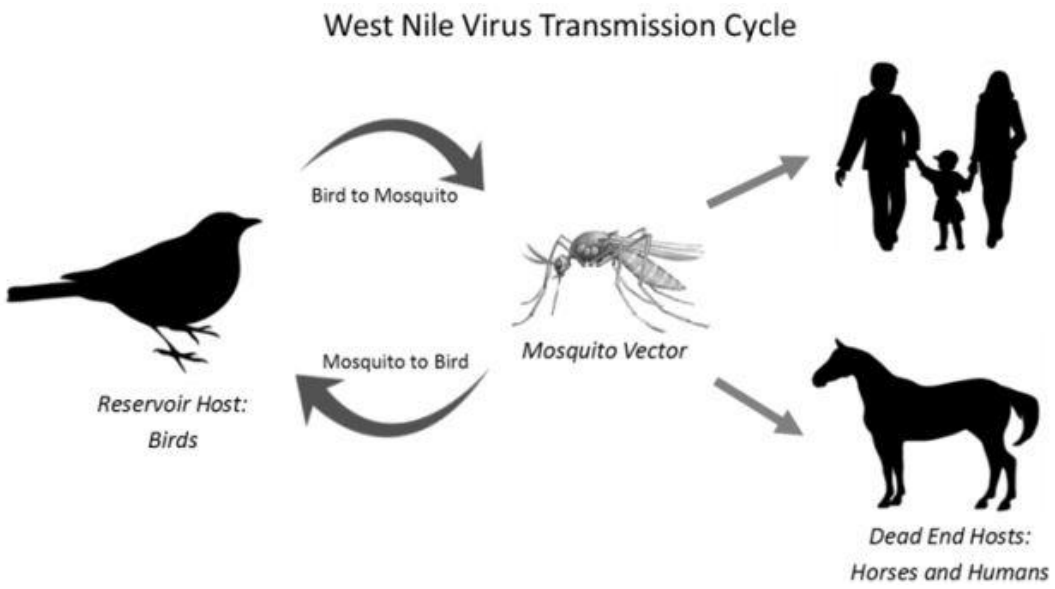


Figure 1. The West Nile Virus Transmission Cycle Including Reservoir and Dead End Hosts

From its fist isolation in 1937, reports of both sporadic and epidemic outbreaks of WNV transmission have been documented in Africa, the Middle East, Europe, and Asia. Historically, symptoms of the virus were reported beginning in the 1950's from Egypt and Israel, and continued through the 1960's and 70's in France and South Africa respectively. Since the early 1990's, outbreaks in Romania, Morocco, Italy, Russia, Israel and North America have been recorded (Dauphin et. al, 2004). After WNV's first appearance in the United States in 1999, it spread across the continent by traveling in a cyclic pattern of mosquito to infected bird. As such, over the course of five years, the virus spread across the contiguous forty-eight states and southern Canada resulting in 16,706 human cases and over 700 fatalities (West Nile Virus, 2013). In 2012, the United States experienced the largest national outbreak of the virus since 2002. In total, 2,873

West Nile neuroinvasive disease cases were reported to the Centers of Disease Control and Prevention (West Nile Virus, 2013). To date, only 117 out of 3,140 US counties nationwide have never detected the virus in humans or animals (Green & Reid, 2013). Presently, WNV has become the most significant mosquito-borne disease in North America and currently has the broadest global geographic distribution of any contemporary vector-borne disease (Hofmeister, 2011).

2.2. *Culex pipiens*

While WNV has been identified in sixty-five mosquito species since 1999 in the United States, the predominant vector in the northeastern and north central US is the *Culex pipiens* mosquito (Andreadis et al., 2004; Hamer et al., 2008; Kilpatrick et al., 2005; Turell et al., 2002; West Nile Virus, 2013). *Cx. pipiens* belongs to the *Cx. pipiens* complex: a group of evolutionarily closely related species that are often difficult to distinguish morphologically (Collins and Paskewitz, 1996). The *Cx. pipiens* complex is comprised of: *Cx. pipiens, Culex quinquefasciatus, Culex australicus,* and *Culex globocoxitus* (Farajollahi et al., 2011).

Adult *Cx. pipiens* are a small to medium sized mosquito with a light brown thorax and darker brown, segmented, abdomen. The proboscis, palps, tarsi, and wings are all characteristically dark (Green & Reid, 2013). Females feed primarily on songbirds, but also will also draw a bloodmeal from humans and other mammals. The species propensity for entering homes in search of blood has earned it the common name, the "Northern House Mosquito." (Burkett-Cadena, 2013). A container breeding species,

larvae can be found in a variety of habitats including marshes, ditches, discarded automobile tires, sewage catch basins, and a multitude of water-filled, artificial containers (Burkett-Cadena, 2013). Females will lay their eggs in rafts that contain 150-350 eggs; eggs typically hatch within two days (R. Harrison, personal communication, September 12, 2013).

The *Cx. pipiens* geographical range spans forty-one of the contiguous forty-eight states and can be found in both rural and urban areas (Ward, 2005). The success of the species can partially be attributed to their exploitation of "food" found in stagnant water generated by humans and livestock. Unlike other mosquito species, the *Cx. pipiens* mosquito commonly thrives in aquatic environments rich in organic content (Vinogradova, 2000). In recent years, it has been hypothesized that the species high abundance in urban environments is a key factor in the annual rise of WNV transmission rates in urbanized areas (Magori, 2011; Brown et al., 2008; Gomez et al., 2008).

2.3 Habitat Suitability Modeling

Habitat suitability modeling serves to produce spatial predictions of the suitability of locations for a focal species and their potential distribution over a given geographical area. As these types of models often aid in better understanding the ecological niche requirements of a species, they are gaining interest in tackling conservation issues and evaluating the risk of exposure to infectious diseases and their vectors including malaria (Rogers et al., 2006; Peterson et al., 2009; Levine et al., 2004), chagas disease (Peterson et al., 2002), and dengue (Benedict et al., 2007). Prediction and modeling of a species

geographical distribution can be accomplished through mathematical models which relate directly to field observations of occurrence and a set of environmental variables (Hirzel et al., 2002; Kirkpatrick M, 1997; Phillips et al., 2006).

As mosquitoes are poikilothermic animals, a dependency exists between these vectors and specific abiotic conditions, namely habitat and sensitivity to variation in temperature and humidity. The length of mosquito genotrophic cycles, along with the developmental rates of eggs, larvae, and pupae are dependent upon temperature and humidity (Madder et al., 1983; Reisen, 1995; Rueda et al., 1990; Vinogradova, 2000). As such, *Cx. pipiens* mosquito population densities have been shown to vary strongly with latitudinal boundary and upper elevations limits as a direct result of temperature, (Chuang et al., 2012; Wang et al., 2011). According to Morris (2003: 2), suitable habitat can be considered, "A spatially bounded subset of physical and biotic conditions among which population density of a focal species varies from adjacent subsets." Accordingly, these subsets are characterized by a combination of abiotic and biotic processes that allow for the distribution and survival of the species (Hutchinson, 1957; Hirzel, 2002).

In recent years, satellite data have increasingly been used to generate risk and habitat suitability maps for disease vectors (Brown et. al 2008; Kitron et. al., 1996), in modeling the geographic distribution of mosquito species in Africa (Rogers et. al., 2000), and in predicting the densities of anopheline vectors (Wood et al., 1991). Applications of GIS and Remote Sensing have additionally served in identifying the correlations between population densities and temperature, humidity's influence on larval development, the

length of the genotrophic cycle, and the extrinsic incubation period (Pope et al., 1992; Wood et al., 1992; Beck et al., 1994; Dister et al., 1997).

The generation of species habitat suitability maps is often accomplished through the combination of continuous surface data of species abundance, derived either through an interpolation or landscape based approach, and multivariate analysis including logistic regression (Peeters and Gardeniers, 1998; Higgens et al., 1999; Manel et al., 2001; Palma et al., 1999), Gaussian logistic regressions (Ter Braak, 1987; Legendre and Legendre, 1998), discriminant analysis (Legendre and Legendre, 1998; Livingston et al., 1990; Manel et al., 1999), Mahalanobis distances (Clark et al., 1993), and artificial neural networks (Manel et al., 1999). A key function of each of these types of analysis includes sampling for presence/absence species data. In presence/absence modeling, each sample site is monitored in order to affirm with sufficient certainty either the presence or absence of the species (Hirzel, 2006). However, absence data can often be difficult to accurately obtain either due to the species lack of detection despite its known presence, or the habitat is truly not suitable for the species. The inclusion of "false absences" within a dataset may result in biased analysis and must be carefully considered (Hirzel, 2002).

In order to combat issues of presence-absence modeling, a number of approaches have been developed for presence-only modeling. The Genetic Algorithm for Rule-Set Prediction (GARP) (Stockwell & Nobel, 1992; Stockwell, 1999) produces a set of positive and negative rules that together output a binary prediction. This method is based on genetic algorithms drawn from an artificial-intelligence framework. In GARP

modeling, the positive and negatively established rules are favored according to their significance based upon a sample of background and presence pixels within the study area (Stockwell, 1999).

Ecological Niche Factor Analysis (ENFA) (Hirzel et al., 2002) is a species distribution model which builds upon Hutchinson's (1957) concept of an ecological niche: A hyper-volume in the multidimensional space of ecological variables within which a species can maintain a viable population. ENFA is similar to Principal Component Analysis as it possesses the ability to summarize data in terms of two orthogonal factors: Marginality and Specialization. Environmental suitability is then modeled as a Manhattan distance in the transformed space (Phillips et. al, 2006).

BIOCLIM (Nix & Busby, 1986) outputs suitable environmental conditions as a "bioclimatic envelope" that represents the overall range, or a percentage of, observed presence values in each of the input environmental dimensions. Similarly, DOMAIN (Carpenter et al., 1993) uses a computed metric where a predicted suitability index is calculated by computing the minimum distance in environmental space to any and/or all of the presence records.

2.4 Maximum Entropy (MaxEnt) Modeling

An additional presence-only modeling technique, MaxEnt is a general purpose machine learning method for modeling a focal species' likely geographic distribution from a set of presence-only occurrence localities and a set of environmental variables. An occurrence locality is defined as a latitude-longitude pair that demarcates a site where the

species has been observed (Phillips et. al, 2006). MaxEnt works by estimating a target probability distribution by finding the probability distribution of maximum entropy (i.e., the most spread out, or closest to uniform), subject to a set of environmental constraints that represent our incomplete knowledge about the target distribution (Phillips et al., 2006). MaxEnt is based upon the maximum-entropy principle as defined by J.T. Jaynes, where the best approach to approximating an unknown probability distribution is ensuring that the approximation will satisfy any existing constraints on the unknown distribution and that subject to the existing constraints, that the distribution should maintain maximum entropy (Jaynes, 1957).

In MaxEnt modeling, the unknown probability distribution, π is over a finite set of pixels, X; the study area. Individual elements of X are referred to as points. The distribution of π assigns a non-negative probability $\pi(x)$ to each point x, where all probabilities sum to 1. The approximation of π is also regarded as a probability distribution and is denoted as $\hat{\pi}$. Entropy of $\hat{\pi}$ is defined as:

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) ln\, \hat{\pi}(x)$$

where ln is the natural logarithm. The entropy is non-negative and is at most the natural log of the number of elements in X when all probabilities are equal (Phillips et. al, 2006). As maximum-entropy involves the amount of choice involved in the selection of an event

(Shannon, 1948), the maximum entropy principal applied to MaxEnt methodology is interpreted as, no unfounded constraints being placed upon $\hat{\pi}$ (Phillips et. al, 2006).

Several assumptions are taken into account when formalizing the constraints on the unknown probability distribution of π. The first assumes that there exists a set of known real-valued functions $f_1 \ldots, f_n$ on *X*, known as "features." Features in MaxEnt are derived from ecogeographical data variables of two types: Continuous and Categorical. Continuous variables take arbitrary, real values which correspond to measured quantities. Categorical variables take only a limited number of discrete values. MaxEnt implements features in six classes: Linear (L), Quadratic (Q), Product (P), Threshold (T), Hinge (H), and Category (C) (Table 1) (Phillips & Dudik, 2008). For example, if Maximum Temperature (TMax) is used as a predictor variable, linear transformation ensures that the mean value of TMax where the species is predicted to occur will approximately match the mean value where it is observed to occur. Quadratic features will constrain the variance in TMax across the species predicted area to match observations. Product features will constrain the covariance of TMax with other predictor variables. Threshold features generate a continuous presence binary by making a feature whose value is 0 below the threshold and 1 above. Hinge features are similar to threshold, but a linear function rather than a step function is applied. All categorical features (e.g. land use) will split a predictor with *n* categories into *n* binary features, which take the value 1 when the feature is present and 0 when it is absent. All features are mathematically rescaled to the

interval [0,1] to allow for comparison between coefficients during modeling (Merow et al., 2013).

Table 1. MaxEnt Feature Types

Feature Class	**Description in relation to ecogeographical variable**	**Constraint imposed on estimated distribution $\hat{\pi}$**
Linear (L)	Variable itself	The mean of variable under $\hat{\pi}$ should be close to its mean in the sample locations
Quadratic (Q)	Square of variable	If used with L, variance of variable under $\hat{\pi}$ is close to its variance in the sample
Product (P)	Product of two variables	If used with linear features for the two variables, that the covariance of the variable under $\hat{\pi}$ should be close to the covariance in the sample
Threshold (T)	A step function that allows a different response below the threshold ("the knot") to that above it. Equivalent to a piecewise constant spline.	The proportion of $\hat{\pi}$ that has values of this variable above the knot should be close to that proportion in the sample
Hinge (H)	Similar to the threshold feature, but the response above the knot or below the knot is linear with a positive or negative coefficient (slope). Equivalent to a piecewise spline.	The mean of the variable above the knot under $\hat{\pi}$ should be close to its mean above the knot in the sample locations
Category (C)	A binary indicator showing membership in one call of a categorical variable. For a k-class categorical variable there will be k categorical features	The proportion of $\hat{\pi}$ that has values in this class should be close to that proportion in the sample

The second assumption when formalizing the constraints on the unknown probability distribution of π, assumes that the information known about π are characterized by the averages of the features under π. Each feature f_j assigns a real value

$f_j(x)$ to each point, *x* within the study area. Thus, the expectation of feature f_j under π is denoted by $\pi[f_j]$ and is defined as:

$$\sum_{x \in X} \pi(x) f_j(x).$$

For any probability distribution *p* and function *f,* the notation *p*[*f*] is used to express the expectation of *f* under *p.* The maximum entropy principle ultimately seeks the probability distribution $\hat{\pi}$ of maximum entropy, subject to the constraint that each feature f_j, has the same mean $\hat{\pi}$ as observed empirically (Phillips et. al, 2006).

Two types of maximum entropy modeling exist: conditional and unconditional. Conditional MaxEnt modeling serves to approximate a joint probability distribution *p*(***x***,*y)* for the inputs ***x*** and output label *y.* In this type of modeling scenario, both presence and absence data on the focal species are required for training purposes. In the unconditional maximum likelihood formulation, MaxEnt probability distribution is calculated first by starting with a uniform probability distribution, for which the vector of *n* real-valued coefficients or feature weights, $\lambda = (0,\ldots 0)$. Then, repeated adjustments are made to one or more weights λ_j in order that regularized log loss is decreased. A deterministic algorithm, the results are guaranteed to converge to the MaxEnt probability distribution. The algorithm will iterate until the change in log loss reaches a user-specified convergence threshold or the maximum number of user-specified iterations have been performed (Phillips et. al, 2006).

CHAPTER III

METHODOLOGY

3.1 Site Selection

The study area for this research was Forsyth County, North Carolina, a 413 square mile area located in the North Central Piedmont. This study area was chosen based upon geographical proximity and a known presence of *Cx. pipiens* larval populations from previous field collections. Forsyth County was also among the seven counties that reported a neuroinvasive case of WNV to the Centers for Disease Control (CDC) during the 2012 trapping season.

The total population for Forsyth County in 2010 was 350,670 with 90.9% of the population residing in urbanized areas. Of the remaining 9.1% of the population, rural-farm and non-farm residents were 0.2% and 8.9% respectively. The city of Winston-Salem is the largest city in the county, with a 2010 population of 229,617 residents. Other larger, more notable cities by population in the county are Kernersville and Clemmons with 2010 populations of 23,123, and 18,627 respectively (US Census Bureau, 2010).

Elevation within the study area ranges from 155 to 336.4 meters (508.53ft to 1103.67ft) above sea level. The county is delineated to the east by a fall line and to the

west by the base of the Appalachian Mountain range. Of the total land area, three square miles (.72%) is water; most notably the 365-acre, Salem Lake.

3.2 Mosquito Sampling

Mosquito trapping data for this study were acquired by Forsyth County Vector Control in sixty-nine locations throughout the study area over a twenty-eight week period during the 2013 collection year (April to October). Sampling sites for this study were not randomly selected. Selection of trapping locations were based upon previous collection of the *Cx. pipiens* species and mosquito control personnel's need to maximize local collection in urbanized areas. The number and spatial distribution of trapping sites varies across the area annually based on funding, personnel, and resident request. Additionally, trapping frequency at each site was irregular; traps were not set at each site on a weekly basis, but may have been set on either a bi-weekly or tri-weekly basis based on vector control personnel needs. Mosquito traps were set from Monday to Thursday and the samples collected the next day (Tuesday to Friday). Collections were performed using gravid traps; a standard tool for mosquito-borne disease surveillance (Figure 2). Captured mosquito species were identified by vector control staff according to the species' morphological characteristics. All collected mosquitoes were sorted by species to allow for viral assay for WNV to be conducted by state health officials.

Figure 2. Gravid Mosquito Trap used for Sampling Collection.

Trapping sites included parks, college campuses, tire dumps, undeveloped wood lots, wetlands, and densely residential locales. Individual collection locations were recorded using a Global Positioning System (GPS) unit in order to provide locational point data for each site and its surrounding areas. All GPS locational point data were imported to the ArcMap environment as a point shapefile and merged to collection documentation including collection date, location ID, total mosquito collection, and collected species (Figure 3). A selection of only those occurrence localities with a positive *Cx. pipiens* presence were then extracted to serve as separate layer of species

presence for model processing. A total of thirty-two sites comprised the positive species localities layer.

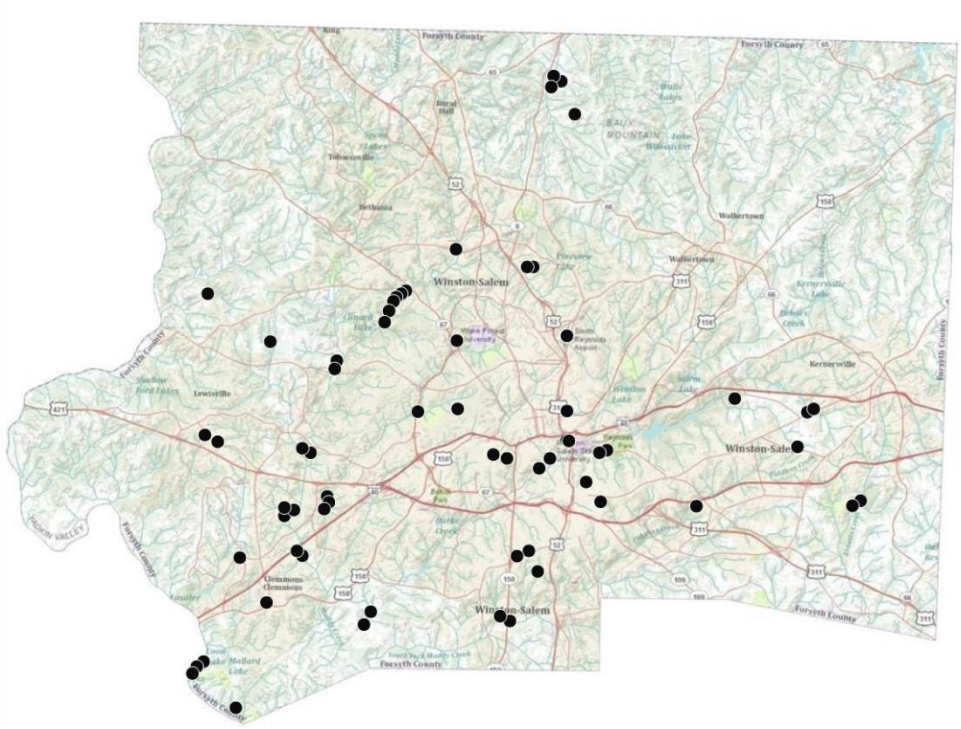


Figure 3. Sampling Locations for *Cx. pipiens* Mosquito Species in Forsyth County, NC

3.3 Ecogeographical Data

Ecogeographical predictor variables relevant to the prediction of the *Cx. pipiens* species were selected based on an assessment of the biological characteristics of the species. Taking into account the size of the study area, data source selections were chosen on the basis of availability and spatial scale. Four categories of environmental descriptors, totaling eleven ecogeographical variables contributed to the development of

the final predictive model. All variables participating in this study were formatted in raster format and processed in ESRI's ArcMap 10.2 environment.

3.3.1 Topographic Variables

Forsyth County belongs to a total of three drainage basins: Yadkin-PeeDee, Cape Fear, and Roanoke. The counties who's DEMs were processed as part of these three drainage basins included: Surry, Stokes, Guilford, Rockingham, Davie, Davidson, Wilkes, Caldwell, Watauga, Yadkin, Carroll, VA, and Patrick, VA. All topographic variables were computed from 10-meter Digital Elevation Models (DEMs) obtained from the National Elevation Dataset (NED). Higher resolution data sources used to derive these imagery products include light detection and ranging (lidar), interferometric synthesis aperture radar (ifsar) and high-resolution imagery. The topographic variables included in this study as calculated from the 10-meter DEMs were elevation, slope, aspect, and a hydrographic network (Figure 4). To accurately calculate the hydrographic network, all basins that drained into as well as those within Forsyth County were used in this model.

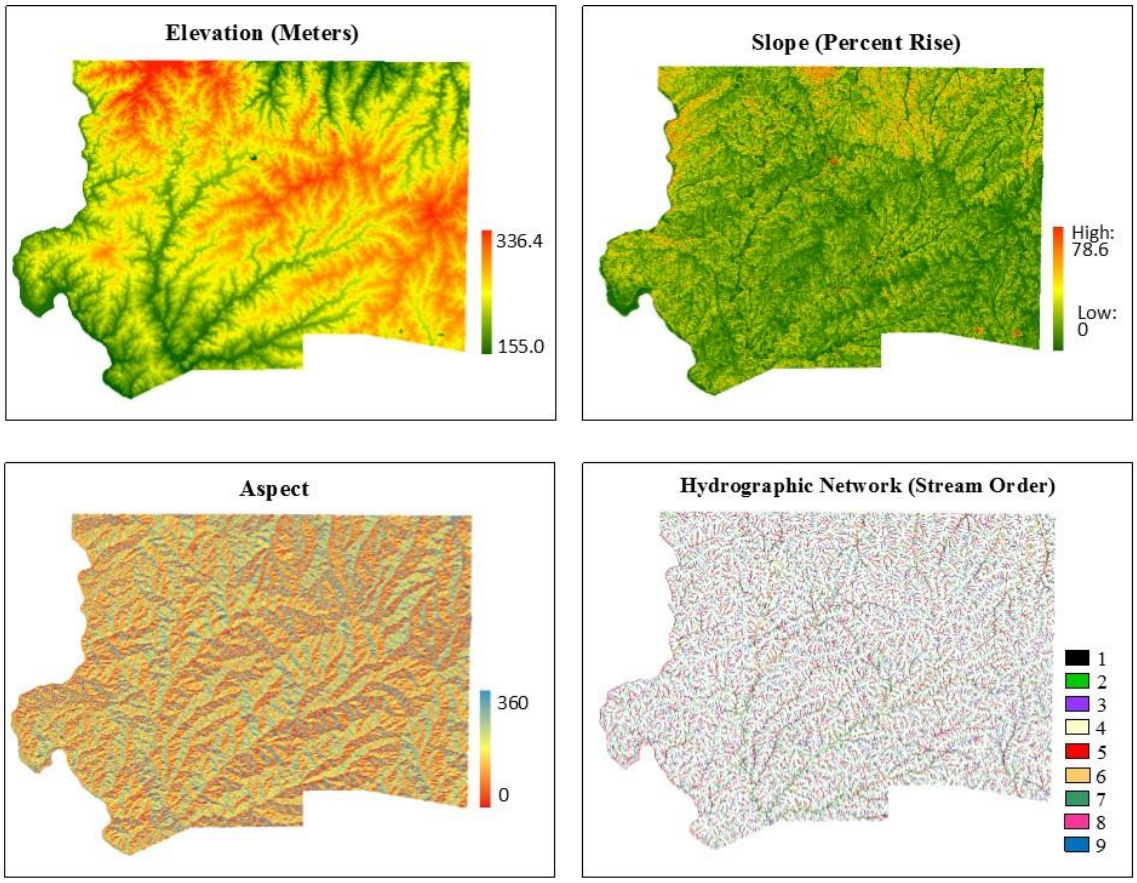


Figure 4. Topographic Variables Generated for Model Processing

3.3.2 Climatic Variables

Climatic variables for this study included temperature minimums (TMin) and maximums (TMax) (in °F), precipitation (in inches), and evapotranspiration. Precipitation, TMin and TMax values were acquired from National Climate Data Center (NCDC) weather stations throughout both the study area and queen case adjacent counties in order to lessen the severity of edge effects on interpolated surfaces. Sixty stations served to interpolate precipitation values and forty-five stations for TMax and TMin (Figure 5); Stations were chosen based on data availability. These data were thirty-

year, monthly normals averaged over the duration of the breeding season (April to October). Interpolation of climate values was conducted using the Ordinary Kriging method: an interpolation method which makes the best use of the values inferred from control point data in order to interpolate an optimal surface structure output (O'Sullivan & Unwin, 2010). Resulting surfaces were rasterized to 10-meter.

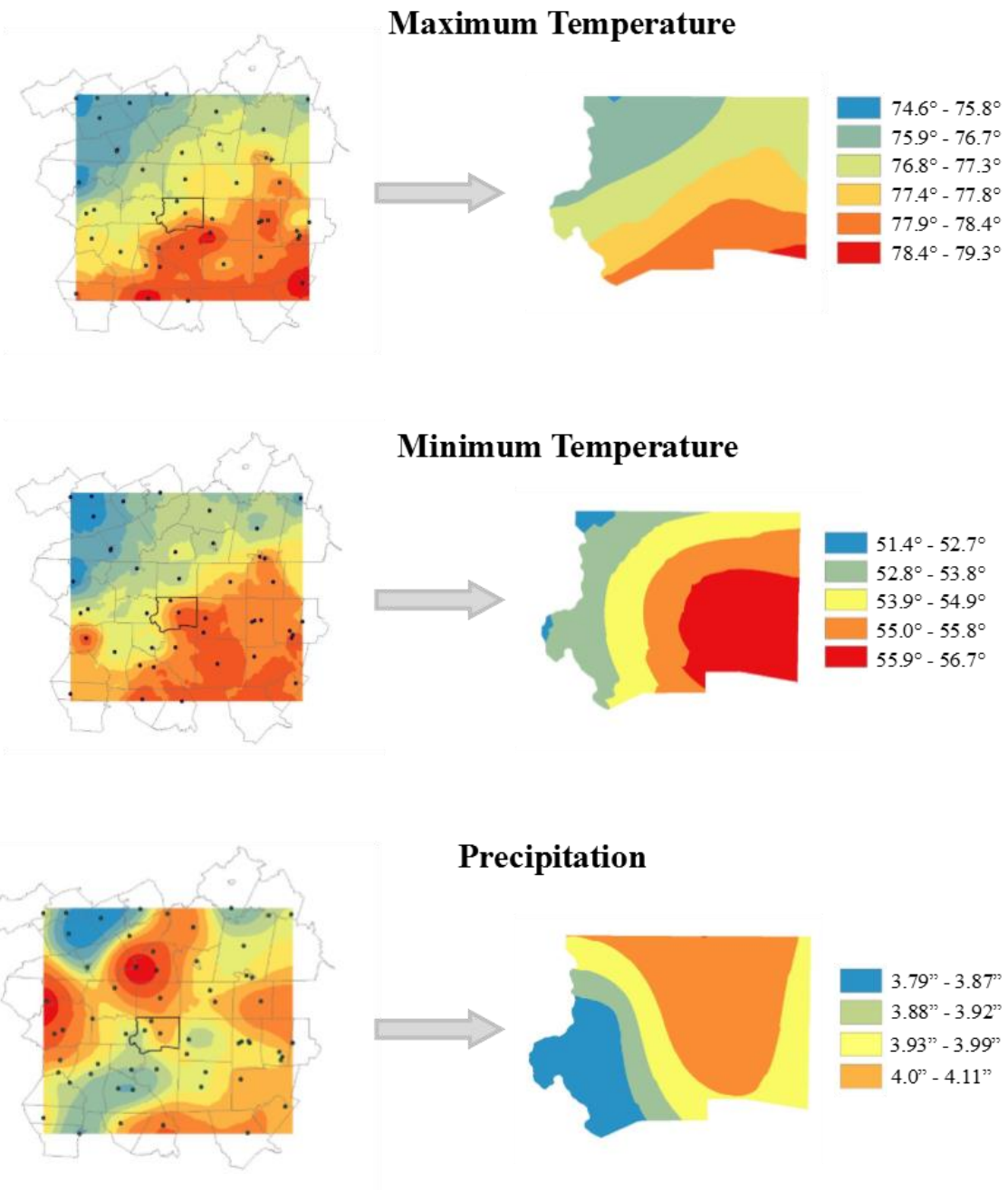


Figure 5. Ordinary Kriging of Maximum Temperature (TMax), Minimum Temperature (TMin), and Precipitation Surfaces

Evapotranspiration data were imported in to the ArcMap environment using the Import Evapotranspiration tool as a part of the MODIS Toolbox developed by Daniel Siegel. The toolbox contains scripts which allow for the importation of historical MODIS imagery data products and is available for download in the ArcGIS Resource Center (http://resources.arcgis.com/gallery/file/geoprocessing/details?entryID=9CC382D2-1422-2418-34F8-DC9F97B24052). Global evapotranspiration data, MOD16, were developed by Dr. Qiaozhen Mu at the University of Montana. Imported evapotranspiration data are calculated using the Penman-Monteith equation using both land surface temperature (MOD11) and albedo (MOD43) datasets. Estimated global evapotranspiration is available for the entire globe at a resolution of 1-kilometer dating back to January 2000 (Siegel) .

3.3.3 Habitat Variables

Land Use data were classified from 2010 imagery derived from the National Agriculture Imagery Program (NAIP). NAIP imagery is acquired at a one-meter ground sampling distance during the agricultural growing season. For the purposes of this study, land use was classified in to six classes: urban, forested, barren, water, agricultural, and shadow. The urban or built-up land classification included two sub-classes: residential and non-residential. United States Geological Survey (USGS) hydrography data along with Forsyth County building footprints, zoning classification, and street centerlines combined with visual inspection assisted in compiling the final classification output (Figure 6).

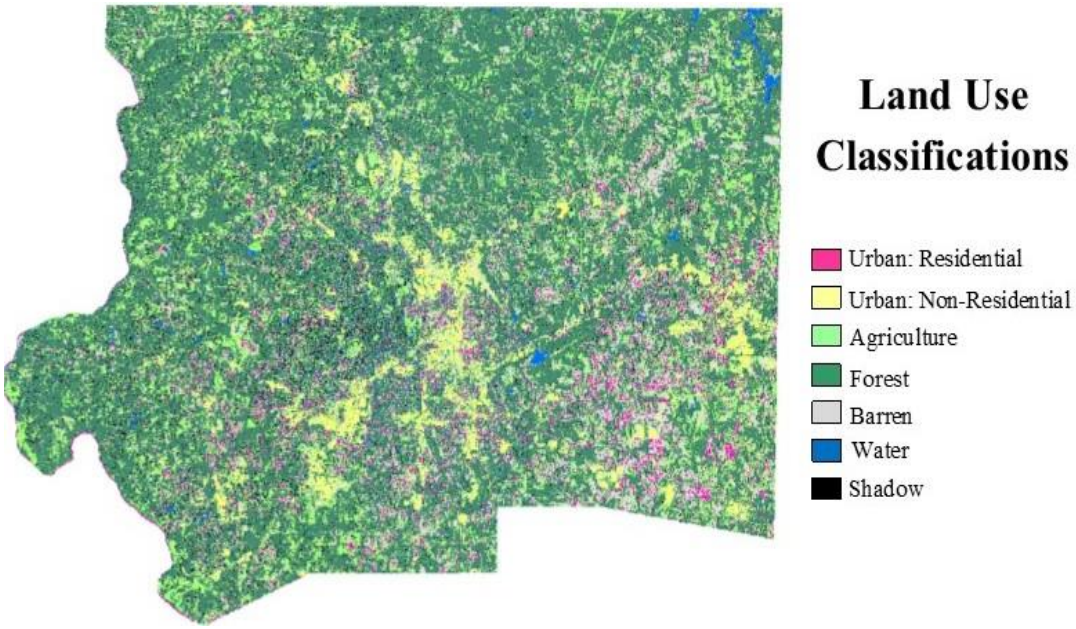


Figure 6. Land Use Classifications as Habitat MaxEnt Variable

3.3.4 Human Population Variables

Considered an "urban" mosquito, *Cx. pipiens* are known to breed in a wide range of areas including domestic sites, sewer catch basins, and in artificial containers. In order to model the association between high human population density and presence of *Cx. Pipiens*, median household income (Figure 7) and population density (Figure 8) layers were generated from 2010 census data on the census tract levels. Previous studies have highlighted a positive correlation between human population density, urban morphology, and the mean number of female *Cx. pipiens* mosquitoes (Andreadis et al., 2004; Tran et al., 2002). Rios et al., (2006), highlights a trend of arbovirus activity (in mosquitoes and humans) in geographical areas associated with socioeconomic status in the local community. Findings suggest that populations residing in virus-positive census tracts

attained less education, maintained a lower median household income, and were more likely to fall below the poverty level (Rios et al., 2006).

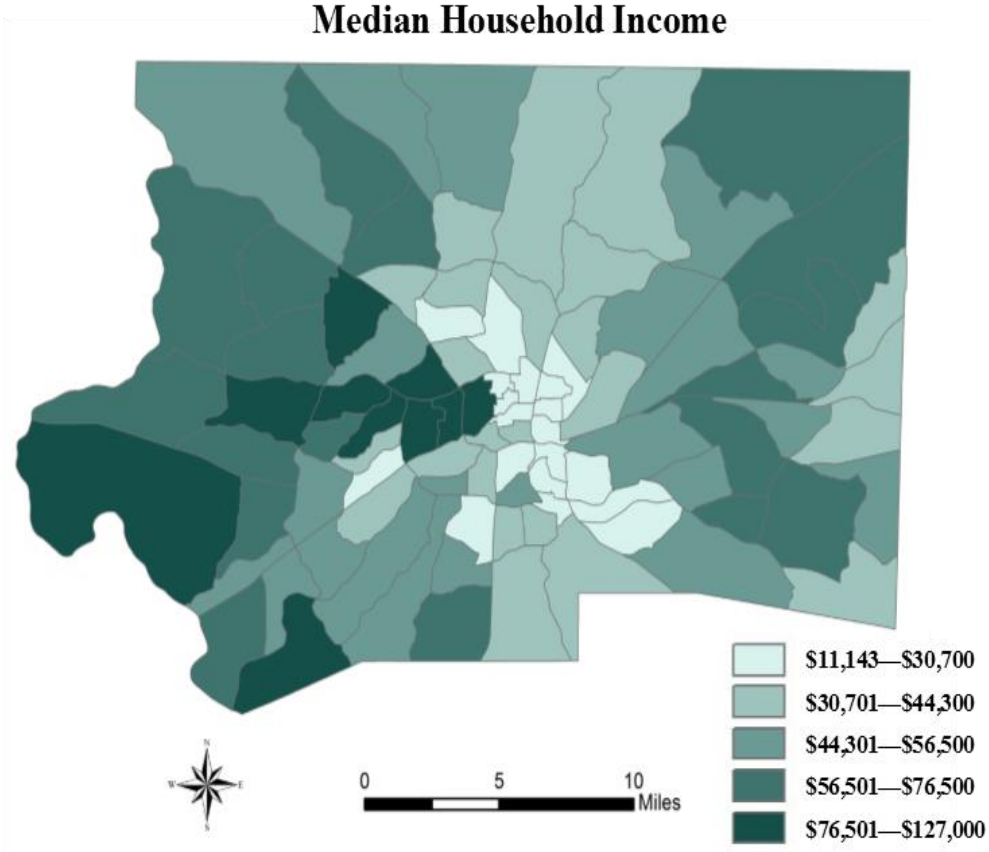


Figure 7. Median Household Income as MaxEnt Variable.

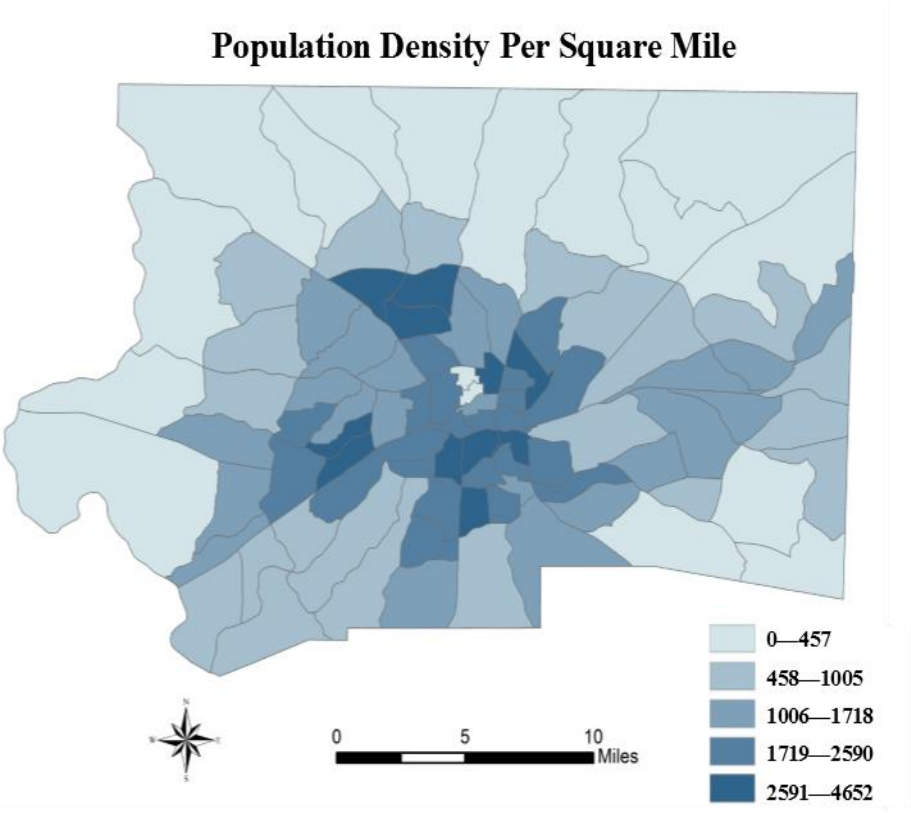


Figure 8. Population Density per Square Mile as MaxEnt Variable.

Table 2. Ecogeographical Variables Used in Modeling Construction.

	Variable	**Data Sources**
Topographic	Elevation Slope Aspect Hydrographic Network	National Elevation Dataset (NED)
Climatic	Temperature (Max) Temperature (Min) Precipitation Evapotranspiration	National Climate Data Center (NCDC) ArcGIS MODIS Toolbox Extension
Human Population	Population Density Median Household Income	US Census Data
Habitat	Land Cover	National Agriculture Imagery Program (NAIP)

3.3.5 Ecogeographical Data Processing

All vector layers participating in this analysis (Positive Species Collection Localities, Hydrographic Network, Median Household Income, Population Density) were rasterized 10m and clipped to the Forsyth boundary extent; thus, ensuring equalization in resolution and spatial extent for the area of interest. Similarly, remaining raster imagery and data layers (DEMs) were normalized to 10m raster's and subset to the county boundary. Equalization to 10m served as a resolution compromise between higher and lower resolution input datasets while maintaining a reasonable processing speed for all data. All data layers were projected using the North American Datum 1983, Universal Transverse Mercator (UTM) Zone 17 North to ensure proper co-referencing of planimetric (x,y) cell locations. Finally, all ecogeographocal data layers were exported from the ArcMap 10.2 environment as ascii (.asc) files for analysis in MaxEnt (Figure 9).

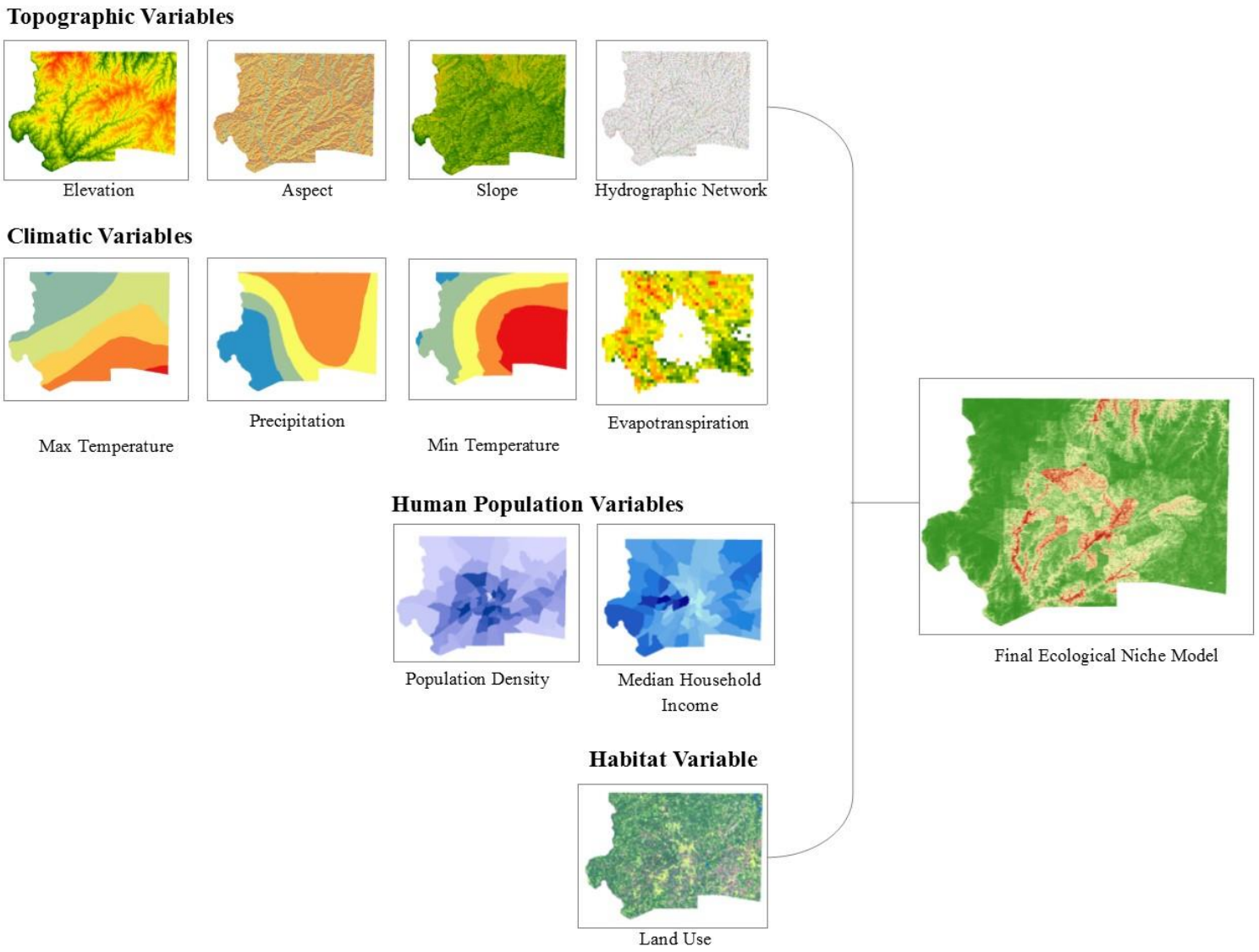


Figure 9. Modeling Workflow

3.4 MaxEnt Modeling Implementation

Models generated in this research were created using Maximum Entropy Species Distribution Modeling, Version 3.3.3k. The MaxEnt software package (Phillips et al., 2006) is a particularly popular species distribution/environmental niche model, with over 1000 applications published since 2006 (Merow et al., 2013). MaxEnt modeling was chosen for the purposes of this study due to its ability to model species habitat based upon presence-only data and environmental information. Furthermore, MaxEnt modeling was chosen due to its outputs being continuous in nature which allowed for fine distinctions to be made between the modeled suitability of localities across the study area.

Model processing began by uploading both ecogeographical and species presence data files in to MaxEnt's graphical user interface (Figure 10). As MaxEnt can incorporate interactions between both categorical and continuous data sets, Median Household Income, Population Distribution, Hydrographic Network, Aspect, and Land Use were set as categorical variables, while all other ecogeographical variables were defined as continuous. Because MaxEnt affords the ability to run a model multiple times and average the results from the generated models, a total of ten replicates were generated for this study. The number of replicates chosen was based upon the number generated in the Phillips (2006) explanatory study for model processing. Data were separated in to two partitions with 25% of the total sample records set aside for external validation. The occurrence records that were set aside for validation were chosen at random by MaxEnt. The remaining 75% of sample records were used in the construction of the MaxEnt

model. Of the three optional output formats, Raw, Cumulative, and Logistic, a Raw output was chosen based on the recommendation of Merow et al., (2013). Raw outputs do not rely on post-processing assumptions and unlike logistic output are not based on a strong assumption of the value of the probability of presence at 'average' presence locations (Merow et al., 2013).

The default number of iterations, 1000, was selected for this study as it provides adequate time for convergence ensuring that the model is less likely to over-predict or under-predict relationships. Finally, the default Regularization coefficient value, 1 was chosen based on its performance across a range of taxonomic groups (Phillips and Dudik, 2008). Regularization reduces over-fitting by ensuring that empirical restraints are not fit too precisely and that the model is penalized in proportion to the magnitude of the coefficients (Merow et al., 2013).

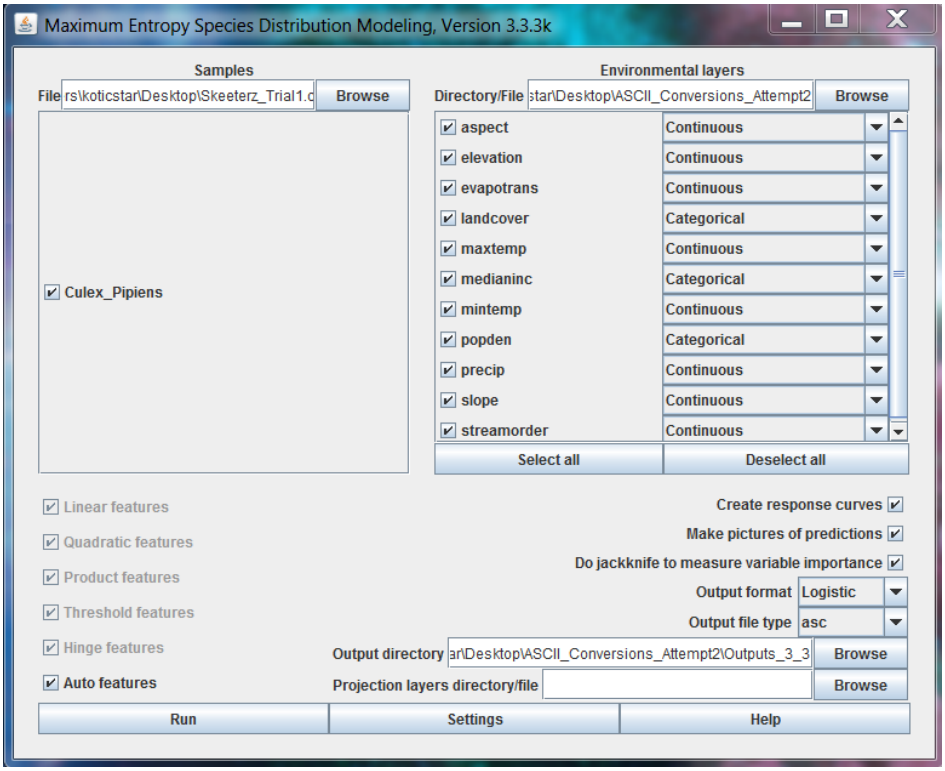


Figure 10. MaxEnt User Interface, Version 3.3.3k.

CHAPTER IV

RESULTS

MaxEnt modeling works by taking a list of user-defined species presence localities as input, as well as a set of ecogeographical predictors across a study area that has been divided in to grid cells. Because the total population size for a given species is typically unknown during modeling, only relative comparisons between species' total population and predicted occurrence rate in each cell are meaningful. This results in a Relative Occurrence Rate (ROR; Fithian and Hastie, 2012), where ROR is the relative probability that any given cell is contained within a collection of presence cells (Merow et al., 2013).

Three probability densities of ecogeographical predictors, **Z** are calculated in MaxEnt's predictions in environmental space: the prior probability density, Q(**z**); the probability density of **Z** at presence locations, P(**z**); and the predicted ROR at each location in the landscape P*(**z**) (Merow et al., 2013). The null hypothesis tested in this study assumed that species were equally likely to be located anywhere within the study area. This assumption meant that every pixel x had the same probability of being selected as background, or equivalently that every environment **z** has a probability of being selected as background according to its frequency P(**z**).

Model evaluation began by performing a threshold-dependent binomial test based on omission and predicted area. Extrinsic omission rates describe the fraction of test locations that fall within pixels that are not suitable for the species. By contrast, the proportional predicted area represents the fraction of all pixels that are predicted as suitable for the species (Phillips et. al, 2006). As 25% of the total sample records were omitted for validation during the initial model run, Figure 11 shows how both validation and training omission versus predicted area vary according to the choice of a cumulative threshold; the graph displays the omission rate and predicted area at different thresholds. Cumulative thresholds assist in determining suitable versus unsuitable habitat. Should a discrete, suitable versus unsuitable habitat model be desired for final modeling output, these values assist in selecting the threshold value that constitutes a suitable habitat. The orange and blue shading surrounding the lines on the graph represent variability.

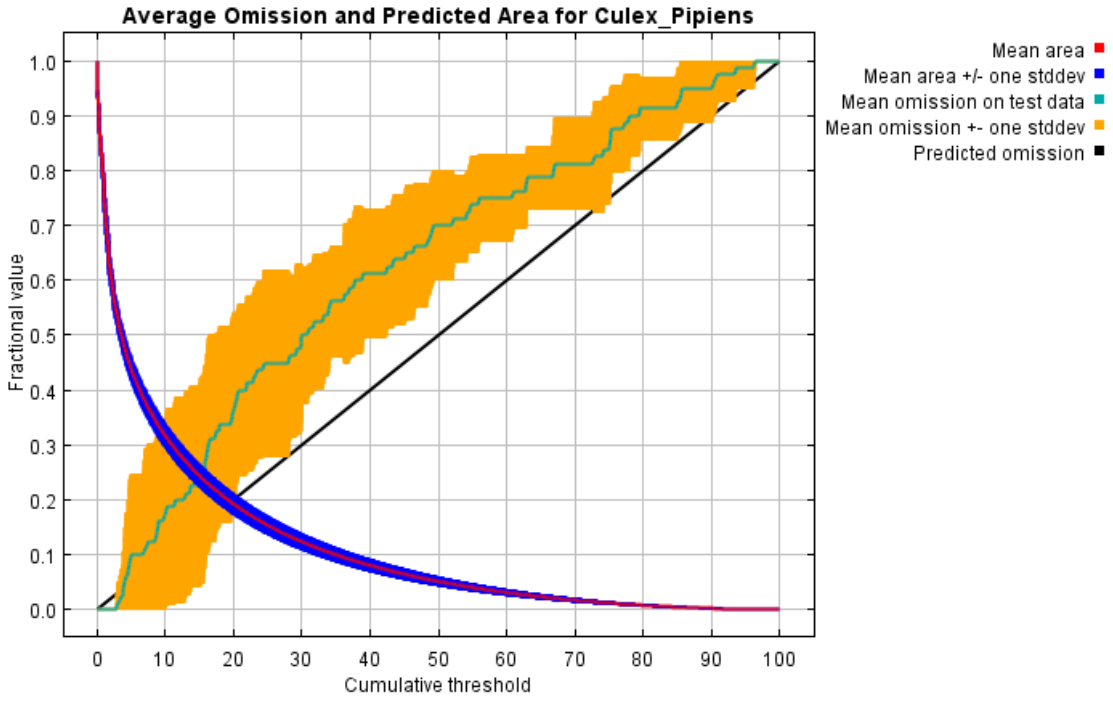


Figure 11. Omission and Predicted Area for *Cx. Pipiens*. The plot depicts how testing and training omission and predicted area vary with choice of cumulative threshold.

Thresholding to produce discrete final models can become problematic as it may depend on species' prevalence or population density, which are typically unknown; thus, it becomes difficult to select a threshold value that is the most biologically meaningful (Merow et al., 2013). According to Merow et al (2013: 1067), "Thresholding is unnecessary in many applications, and embracing the continuous and probabilistic nature of predictions avoids undue confidence in predictions. Often threshold predictions reflect researcher's assumptions about appropriate threshold values and not attributes of the species distribution." Bearing this in mind, the selection of a threshold value to define a discrete suitability model was not employed in this study.

The second approach to model evaluation was using a Receiver Operating Characteristic (ROC) curve (Fielding and Bell, 1997). The area under the ROC curve (AUC) is often used as a single threshold-independent measure for model performance (Manel et al., 2001; Thuiller, 2003; Brotons et al., 2004; McPherson, Ketz and Rogers, 2004; Thullier, Lavorel and Araujo, 2005). An AUC score of one would mean perfect prediction with zero omission. An AUC value lower than 0.5 would indicate that performance of the model is no better than random. The construction of ROC curves is accomplished by using all possible thresholds in order to classify the scores into confusion matrices; obtaining sensitivity and specificity for each matrix. Sensitivity is then plotted against the corresponding proportion of false positives (equal to 1-specificity) (Allouche et al., 2006). Sensitivity is the proportion of observed presences; specificity is the proportion of observed absences. The outcome of this analysis reveals the fit of the model to the training data and the comparison of both validation and training data to random prediction. The AUC values allow for comparison of model performance between model replicates. Threshold evaluation as conducted according to ROC for this analysis revealed that the final, averaged model performed significantly better than random prediction: 0.838 (Figure 12).

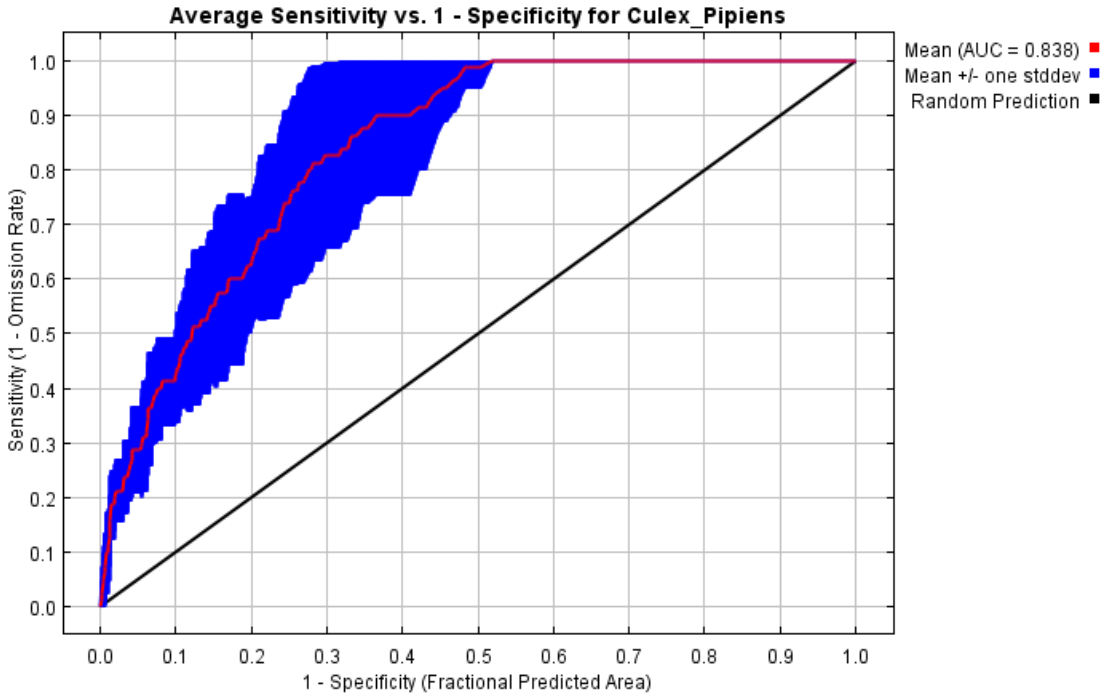


Figure 12. Receiver Operating Curve (ROC) for both Training and Test Data for *Cx. Pipiens*.

Relative contributions of each ecogeographical variable contributing to the final model are summarized in Table 2. In order to determine the final estimate of each variable, at each step in the MaxEnt algorithm, the gain of the model is increased by modifying the coefficient for each feature. Gain is defined as the average log probability of the presence samples included in the study, minus a constant that makes the uniform distribution have zero gain. Gain is a penalized maximum likelihood function where exponentiating the gain function gives the likelihood ratio of an average presence to an average background point. Another words, maximizing the gain will correspond to finding a model that can best differentiate presences from background locations (Merow et al., 2013).

The program assigns the gain increase to the ecogeographical variable(s) that the feature depends on; the conversion of these values to percentages are highlighted in Table 3 under the section heading: Percentage Contribution (Phillips et. al, 2006). All percent contributions are heuristically defined as they depend on a particular path that the MaxEnt code used to reach an optimal solution. The higher the percentage contribution, the more impact that particular variable has on predicting the occurrence of the species. Calculation of percent contribution of each ecogeographical variable reveals that Population Density contributes the greatest amount (25.7%) followed by Maximum Temperature (15.5%) and Elevation (14.7%).

Permutation importance of variables is determined by randomly permuting the values of that particular variable among all presence and background training points and measuring the resulting decrease in training AUC (Phillips et. al, 2006). The contribution of each variable is measured only on the results of the final model, not the paths used to obtain it. A large decrease in permutation for a given variable is indicative of the model depending heavily upon that variable. Elevation (35.3%), Maximum Temperature (22%), and Median Household Income (9.7%) were shown to have the highest overall values of permutation importance of all ecogeographical variables.

Table 3. MaxEnt Analysis of Variable Contribution

Variable	**Percent Contribution**	**Permutation Importance**
Elevation	14.7	35.3
Slope	2	1.5
Aspect	3.8	5.6
Hydrographic Network	0.4	0.2
Minimum Temperature (TMin)	9.1	9.2
Maximum Temperature (TMax)	15.5	22
Precipitation	0.2	0.4
Evapotranspiration	5.3	1.7
Land Use	10.6	5.5
Population Density	25.7	9.3
Median Household Income	12.9	9.7

Additional estimates of variable importance are reported according to jackknifing procedures ran on all ecogeographical variables participating in this study. As the model executes, each variable is excluded in turn, and a model is generated with the remaining variables. Models containing each variable in isolation, as well as a model containing all variables are additionally generated. Two resulting plots assist in ascertaining the value of each ecogeographical participating in the study: Jackknife of Regularized Training Gain for *Cx. Pipiens* and Jackknife of AUC for *Cx. Pipiens*. As noted by the results of the jackknifing procedure of regularized training gain (Figure 13), both Stream Order (Hydrographic Network) and Evapotranspiration achieve very little gain and are therefore, not useful on their own for estimating distributions of *Cx. Pipiens*. By contrast, Population Density and Median Household Income achieve the largest gain and can serve as better individual estimates of species distribution. No variables contain a substantial

amount of useful information that is not already present in other variables, noted by the lack of substantial decrease in the training gain without any individual variables.

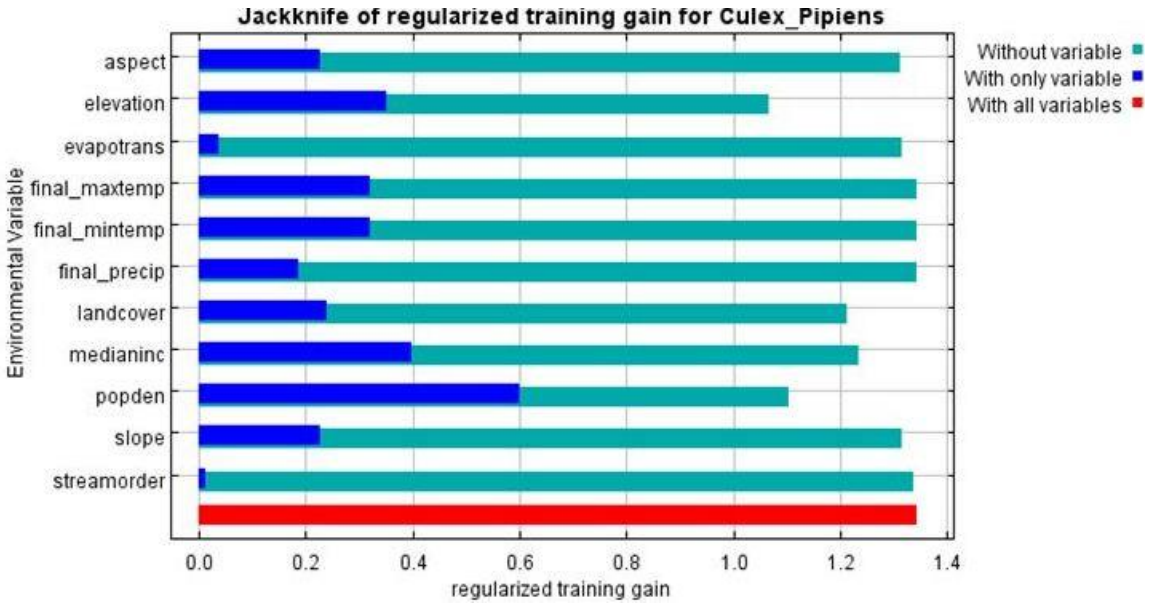


Figure 13. Jackknife of Regularized Training Gain for *Cx. pipiens*

MaxEnt modeling results for the predicted probability of *Cx. pipien*s geographical distribution in Forsyth County highlighted the largest concentrations of *Cx. Pipiens* habitats within and along the periphery of the Winston-Salem municipality (Figure 14). Higher population densities and access to artificial containers for oviposition are readily available within this geographical area; this model is in accordance with the urban nature of the mosquito species. Secondary areas of higher probability are located in the north central portion of the county, an area marked by irrigated cropland and deciduous forest. While the *Cx. pipiens* species is characteristically noted for breeding in urban settings,

the species is known for ovipositing within rural settings as well, though less frequently (Ward, 2005).

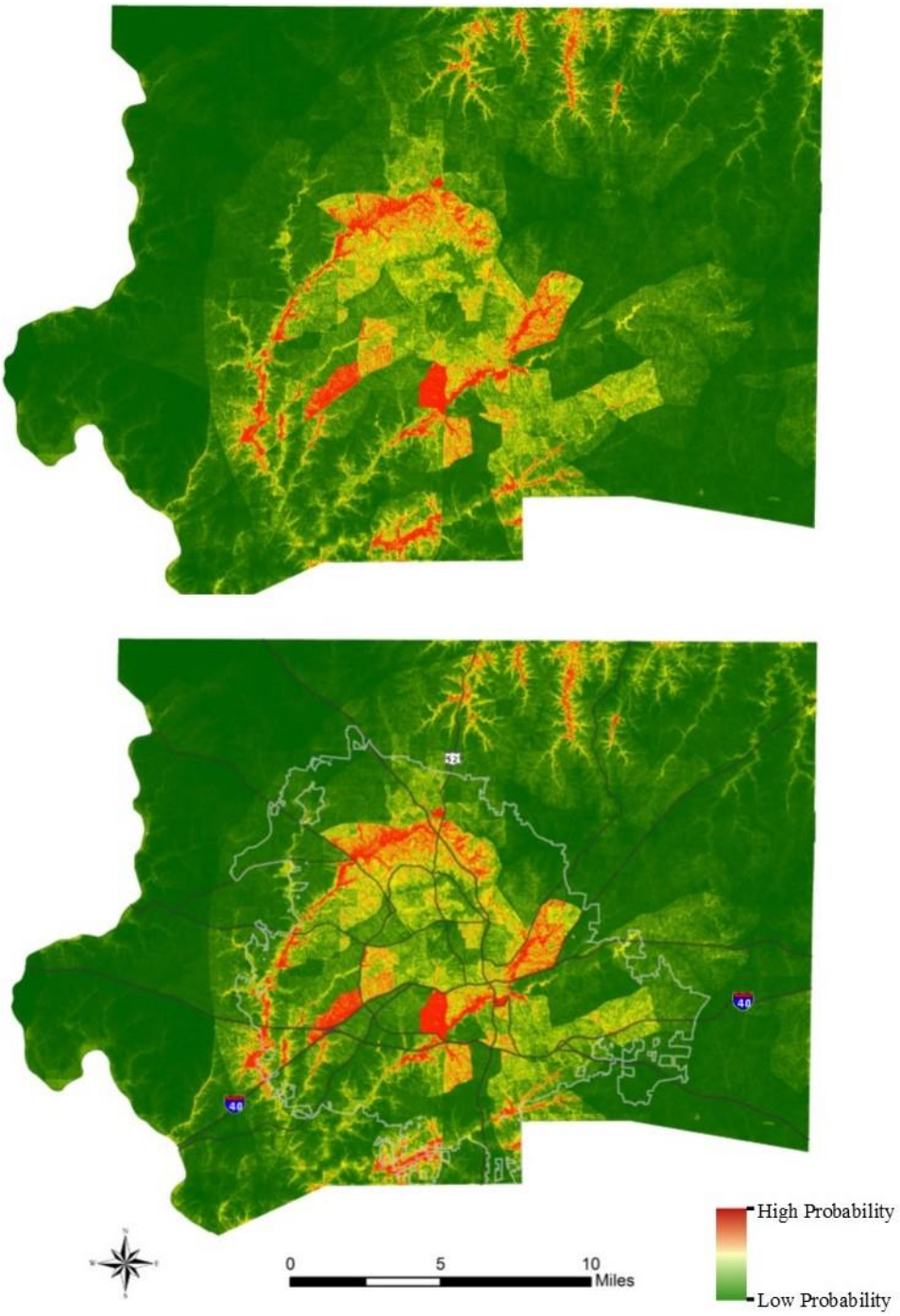


Figure 14. Predicted Probability of *Cx. pipiens* in Forsyth County, NC. Ecological Niche Model output sans boundaries (Top). Modeling output with roads and Winston-Salem municipality highlighted in white.

CHAPTER IV

DISCUSSION

Modeling results highlighted the greatest probability for *Cx. Pipiens* habitats within the Winston-Salem municipality and along the north-central region of the county. These results are driven by the relative importance of Population Density as a predictor variable as indicated by Jackknifing procedures of regularized training gain. The biological nature and characteristics of the *Cx. Pipiens* species allows the mosquito to thrive in urban environments where only small amounts of water and organic matter are necessary for oviposition, and where oviposition can occur in a range of artificial containers. The results of this study further support the urban nature of the species. However, having noted the significance of such variables as Population Density and Median Household Income, which were aggregated to census tract level data, further studies should seek to potentially interpolate these variable surfaces prior to model execution in order to mitigate the Modifiable Area Unit Problems (MAUP) notable in the results of this study as indicated by many linear boundaries separating areas of high and low probabilities. Furthermore, additional studies may consider the use of remotely sensed data as an estimator of human population values such as LandScan data, which at a 1-kilometer spatial resolution is the finest global population distribution dataset available for ambient population rather than census datasets.

All variables modeled in this study were equalized to 10-meter resolution. While many input variables were acquired at this resolution and thus required no further aggregation or resampling, TMax, TMin, evapotranspiration, precipitation, and land cover each required spatial adjustment in order to conform to the 10-meter resolution requirements for modeling. While the potential exists that the spatial adjustment for variables such at evapotranspiration which was resampled from 1-kilometer to 10-meters, may have impacted the variables overall modeling importance, financial restraints, impacts of degradation of spatial resolution on other variables, and model processing time were considered when selecting the modeling resolution for this study. Future studies modeling variables of this nature may seek to acquire datasets with more similar spatial resolutions in order to better understand if these variables importance may have been obscured due to their resolution in this study.

Interpolated surfaces for TMax, TMin, and Precipitation were calculated using thirty year averages for breeding season months (April to October) in order to produce a temporally independent model that would transcend the weather patterns observed strictly during the 2013 breeding season. As the results of this study will serve to assist in trapping placement across the study area for future breeding seasons, climatic variability was an important consideration. Therefore, the use of thirty-year, normal weather data seemed the most appropriate for this study in desiring a model whose results could be used to guide trap placement long term.

Mosquito sampling in this study was conducted entirely by Forsyth County Vector Control where trapping sites were selected based upon previous collection of *Cx. pipiens*. The overarching goal of Vector Control is to maximize the number of collections in urban areas in order to mitigate future viral transmission to the human population. As such, the number and spatial distribution of traps was not randomly selected for this study, but rather was based upon funding, personnel, and resident request. Future studies should seek to randomize trapping placement so as to gain a better understanding of the potential species activity in presently under-sampled rural areas and to reduce potential sampling bias. Furthermore, in planning trap placement, attention should be given to attempting to reduce spatial correlation whenever possible regardless of the scale of future studies. Finally, trapping frequency should be equalized at all sites in order to provide a robust dataset that would include both presence and absence data of the species should an alternative modeling technique be desired.

Prior to MaxEnt modeling, no assessment of autocorrelation between ecogeographical variables was conducted for this study. As a machine learning method, Phillips et al., (2006) and Elith et al., (2011) have noted that high collinearity in predictor variables is less of a problem as compared to standard statistical methods. Thus, including all reasonable predictor variables and allowing the algorithm to select those of greater importance is encouraged. Yet, an alternative school of thought as proposed by Merow et al (2013) and Renner and Warton (2012), suggests minimizing the correlation between ecogeographical variables through conducting correlation analysis, clustering algorithms, principal component analysis, or a similar dimension reduction method prior

to modeling execution. By prescreening variables prior to modeling for autocorrelation, it is argued that modeling will yield a more parsimonious and interpretable overall model. Future work may consider comparing the results of this study having included all variables and an additional study that seeks to remove highly correlated variables in order to assess modeling outcomes.

MaxEnt modeling contrasts species presence against background locations where presence/absence is unmeasured. As such, modifying the background sample for a study is equivalent to modifying the prior expectations for the species distribution. By default, the null hypothesis tested by MaxEnt states that the species is equally likely to be present anywhere on the landscape. By setting the background sample to the entirety of the study area in modeling species distributions at larger geographical scales than was modeling in this study (e.g. State or Country scale), despite limitations that may be present on the species' range (large water bodies, barren land for a species requiring forest canopy, etc.), the number and outcome of potential distribution localities may be altered by MaxEnt. As different background samples can directly impact relative occurrence rate, it is encouraged when modeling at larger geographical scales to consider an ecological justification for background selection.

The comparison of the predicted probability of *Cx. pipiens* habitats across the study area produced in this research and the realized species distribution will serve as the basis for a future study. Trap placement for mosquito sampling will be conducted in

areas demarcated as those areas of both highest and lowest probability distribution for the *Cx. Pipiens* species in order to validate the modeling results derived from this study.

CHAPTER VI

CONCLUSIONS

Modeling results for this study indicated the greatest probability for *Cx. Pipiens* habitats within the Winston-Salem municipality and along the north-central region of Forsyth County. Estimates of variable importance as reported by MaxEnt revealed that Population Density and Median Household Income predictor variables served as the best individual estimates for species distribution. The relationship between the *Cx. pipiens* mosquito and human population is unique as human behavior can largely impact the overall number of potential breeding sites for the species due to the trace amounts of water and organic material necessary for oviposition.

While West Nile Virus (WNV) has been detected in sixty-five mosquito species since 1999 in the United States, the predominant vector in the northeastern and north central US is the *Cx. pipiens* mosquito. Overall, transmission risk depends on vector presence, the productivity of their breeding sites, and location to human settlements and on their effective dispersal. WNV will likely continue to be a public health concern for the foreseeable future due to its establishment in a broad range of ecological settings and transmission through a variety of mosquito species (Hayes et al., 2005). The ability to locate larval habitats and understand their distribution is a critical factor in controlling the abundance of West Nile vectors including *Cx. pipiens*, and mitigating transmission potential to humans. Furthermore, determining geographic areas of higher risk for WNV

in combination with research into new methods to reduce human exposure to mosquitoes may serve to lessen the overall potential severity of WNV transmission.

REFERENCES

Allouche, O.M.R.I., Tsoar, A.S.A.F, & Kadmon, R. (December 01, 2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology, 43*, 6, 1223-1232.

Andreadis, T. G., Anderson, J. F., Vossbrinck, C. R., & Main, A. J. (January 01, 2004). Epidemiology of West Nile Virus in Connecticut: A Five-Year Analysis of Mosquito Data 1999-2003. *Vector Borne and Zoonotic Diseases, 4*, 4, 360-378.

Beck, L. R., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Rejmankova, E., Ulloa, A., ... & Spanner, M. A. (1994). Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. *The American Journal of Tropical Medicine and Hygiene*, *51*(3), 271-280.

Beck, L. R., Rodriguez, M. H., Dister, S. W., Rodriguez, A. D., Washino, R. K., Roberts, D. R., & Spanner, M. A. (1997). Assessment of a remote sensing-based model for predicting malaria transmission risk in villages of Chiapas, Mexico. *The American Journal of Tropical Medicine and Hygiene*, *56*(1), 99-106.

Becker, N., Petric, D., Zgomba, M., Boase, C., Dahl, C., Lane, J., et al (2003). Mosquitoes and their control. New York: Kluwar Academic/Plenum Publisher.

Becker, N., Petric, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., & Kaiser, A. (2010). *Mosquitoes and Their Control*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Benedict, M. Q., Levine, R. S., Hawley, W. A., & Lounibos, L. P. (2007). Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus.*Vector-Borne and Zoonotic Diseases, 7*(1), 76-85.

Bolling, B. G., Barker, C. M., Moore, C. G., Pape, W. J., & Eisen, L. (2009). Seasonal patterns for entomological measures of risk for exposure to Culex vectors and West Nile virus in relation to human disease cases in northeastern Colorado. *Journal of medical entomology, 46*(6), 1519-1531.

Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography, 27*(4), 437-448.

Brown, H. E., Diuk-Wasser, M. A., Guan, Y., Caskey, S., & Fish, D. (2008). Comparison of three satellite sensors at three spatial scales to predict larval mosquito presence in Connecticut wetlands. *Remote Sensing of Environment*,*112*(5), 2301-2308.

Brown, J. H., Mehlman, D. W., & Stevens, G. C. (1995). Spatial variation in abundance. *Ecology*, *76*(7), 2028-2043.

Brownstein, J. S., Rosen, H., Purdy, D., Miller, J. R., Merlino, M., Mostashari, F., & Fish, D. (2002). Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector Borne and Zoonotic Diseases*,*2*(3), 157-164.

Burkett-Cadena, N. D. (2013). *Mosquitoes of the southeastern United States*. University of Alabama Press.

Carpenter, G., Gillison, A. N., & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity & Conservation*, *2*(6), 667-680.

Chuang, T. W., Henebry, G. M., Kimball, J. S., VanRoekel-Patton, D. L., Hildreth, M. B., & Wimberly, M. C. (2012). Satellite microwave remote sensing for environmental modeling of mosquito population dynamics. *Remote sensing of environment*, *125*, 147-156.

Chuang, T. W., Ionides, E. L., Knepper, R. G., Stanuszek, W. W., Walker, E. D., & Wilson, M. L. (2012). Cross-correlation map analyses show weather variation influences on mosquito abundance patterns in Saginaw County, Michigan, 1989-2005. *Journal of medical entomology*, *49*(4), 851-858.

Clark, J. D., Dunn, J. E., & Smith, K. G. (1993). A multivariate model of female black bear habitat use for a geographic information system. *The Journal of wildlife management*, 519-526.

Collins, F. H., & Paskewitz, S. M. (January 01, 1996). A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic Anopheles species. *Insect Molecular Biology*,*5*, 1, 1-9.

Dauphin, G., Zientara, S., Zeller, H., & Murgue, B. (2004). West Nile: worldwide current situation in animals and humans. *Comparative immunology, microbiology and infectious diseases*, *27*(5), 343-355.

Dister, S. W., Fish, D., Bros, S. M., Frank, D. H., & Wood, B. L. (1997). Landscape characterization of peridomestic risk for Lyme disease using satellite imagery. *The American journal of tropical medicine and hygiene*, *57*(6), 687-692.

Diuk-Wasser, M. A., Bagayoko, M., Sogoba, N., Dolo, G., Toure, M. B., Traore, S. F., & Taylor, C. E. (2004). Mapping rice field anopheline breeding habitats in Mali, West Africa, using Landsat ETM+ sensor data. *International Journal of Remote Sensing*, *25*(2), 359-376.

Ebel, G. D., Rochlin, I., Longacker, J., & Kramer, L. D. (2005). Culex restuans (Diptera: Culicidae) relative abundance and vector competence for West Nile virus. *Journal of medical entomology*, *42*(5), 838-843.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43-57.

Farajollahi, A., Fonseca, D. M., Kramer, L. D., & Marm, K. A. (October 01, 2011). "Bird biting" mosquitoes and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology. *Infection, Genetics and Evolution, 11,* 7, 1577-1585.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, *24*(01), 38-49.

Fithian, W., & Hastie, T. (2012). Statistical models for presence-only data: finite-sample equivalence and addressing observer bias. *arXiv preprint arXiv:1207.6950*.

Gomez, A. (January 01, 2008). Land Use and West Nile Virus Seroprevalence in Wild Mammals. *Emerging Infectious Diseases, 14,* 6, 962-965.

Green, S.E., and Reid A. (2013). FAQ: West Nile Virus. Washington, DC: American Society for Microbiology. Retrieved from: http://academy.asm.org/index.php/faq-series/793-faq-west-nile-virus-july-2013?utm_source=mw&utm_medium=ads&utm_campaign=WNV

Guerra, M., Walker, E., Jones, C., Paskewitz, S., Cortinas, M. R., Stancil, A., ... & Kitron, U. (2002). Predicting the Risk of Lyme Disease: Habitat Suitability for *Ixodes scapularis* in the North Central United States.

Guerra, M. A., Walker, E. D., & Kitron, U. (2001). Canine surveillance system for Lyme borreliosis in Wisconsin and northern Illinois: geographic distribution and risk factor analysis. *The American journal of tropical medicine and hygiene*,*65*(5), 546-552.

Hamer, G. L., Walker, E. D., Brawn, J. D., Loss, S. R., Ruiz, M. O., Goldberg, T. L., Schotthoefer, A. M., ... Kitron, U. D. (January 01, 2008). Rapid Amplification of West Nile Virus The Role of Hatch-Year Birds. *Vector Borne and Zoonotic Diseases, 8*, 1, 57-68.

Hayes, E. B., Komar, N., Nasci, R. S., Montgomery, S. P., O'Leary, D. R., & Campbell, G. L. (2005). Epidemiology and Transmission Dynamics of West Nile Virus. *Emerging infectious diseases, 11*(8).

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal*,*327*(7414), 557.

Hirzel, A. H., Hausser, J., Chessel, D., & Perrin, N. (2002). Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data?.*Ecology*, *83*(7), 2027-2036.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *ecological modelling*, *199*(2), 142-152.

Hofmeister, E. K. (January 01, 2011). West Nile virus: North American experience. *Integrative Zoology, 6*, 3, 279-89.

Hutchinson, G. E. (1957). Cold Spring Harbor Symposium on Quantitative Biology. *Concluding remarks, 22*, 415-427.

Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620–630.

Kilpatrick, A. M., L. D. Kramer, S. R. Campbell, E. O. Alleyne, A. P. Dobson, and P. Daszak. 2005. West Nile virus risk assessment and the bridge vector paradigm. Emerging Infectious Diseases 11:425-429

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a species' range. *The American Naturalist*, *150*(1), 1-23.

Kitron, U., Otieno, L. H., Hungerford, L. L., Odulaja, A., Brigham, W. U., Okello, O. O., ... & Cook, E. (1996). Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS.*Journal of Animal Ecology*, *65*(3), 371-380.

Lebl, K., Brugger, K., & Rubel, F. (2013). Predicting Culex pipiens/restuans population dynamics by interval lagged weather data. *Parasit Vectors*, *6*, 129.

Legendre, P., & Legendre, L. (1998). Numerical ecology: second English edition. *Developments in environmental modelling*, *20*.

Levine, R. S., Peterson, A. T., & Benedict, M. Q. (2004). Distribution of members of Anopheles quadrimaculatus Say sl (Diptera: Culicidae) and implications for their roles in malaria transmission in the United States. *Journal of medical entomology*, *41*(4), 607-613.

Livingstone, D. J., Hesketh, G., & Clayworth, D. (1991). Novel method for the display of multivariate data using neural networks. *Journal of molecular graphics*, *9*(2), 115-118.

Madder, D. J., Surgeoner, G. A., & Helson, B. V. (1983). Number of generations, egg production, and developmental time of Culex pipiens and Culex restuans (Diptera: Culicidae) in southern Ontario. *Journal of medical entomology*, *20*(3), 275-287.

Magori, K., Bowden, S., Drake, J. M., & Bajwa, W. I. (July 01, 2011). Decelerating spread of west nile virus by percolation in a heterogeneous urban landscape. *Plos Computational Biology*, *7*, 7.)

Manel, S., Dias, J. M., & Ormerod, S. J. (1999). Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*,*120*(2), 337-347.

Manel, S., Williams, H. C., & Ormerod, S. J. (2001). Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of applied Ecology*, *38*(5), 921-931.

McPherson, J. A. N. A., Jetz, W., & Rogers, D. J. (2004). The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?. *Journal of applied ecology*, *41*(5), 811-823.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058-1069.

Morris, D. W. (January 01, 2003). Toward an ecological synthesis: a case for habitat selection. *Oecologia*, *136*, 1, 1-13.

Nash, D., Mostashari, F., Fine, A., Miller, J., O'Leary, D., Murray, K., ... & Layton, M. (2001). The outbreak of West Nile virus infection in the New York City area in 1999. *New England Journal of Medicine*, *344*(24), 1807-1814.

N.C. Department of Health and Human Services, Mosquito-Borne Illness in North Carolina:West Nile Virus, LaCrosse and EEE. (2013). Retrieved from: http://epi.publichealth.nc.gov/cd/arbo/arbo_fs.pdf

Nix, H. A., & Busby, J. (1986). BIOCLIM, a bioclimatic analysis and prediction system. *Annual report CSIRO. CSIRO Division of Water and Land Resources, Canberra*.

O'Sullivan, D., & Unwin, D. J. (2003). *Geographic Information Analysis*. John Wiley & Sons.

Palma, L., Beja, P., & Rodrigues, M. (1999). The use of sighting data to analyse Iberian lynx habitat and distribution. *Journal of Applied Ecology, 36*(5), 812-824.

Peeters, E. T. H. M., & Gardeniers, J. J. P. 1998. Logistic regression as a tool for defining habitat requirements of two common gammarids. *Freshwater Biol,39*, 605-615.

Peterson, A. T., Sánchez-Cordero, V., Beard, C. B., & Ramsey, J. M. (2002). Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico.*Emerging infectious diseases*, *8*, 662-667.

Peterson, A. T. (January 01, 2009). Shifting suitability for malaria vectors across Africa with warming climates. *Bmc Infectious Diseases, 9*.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling, 190*(3), 231-259.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography, 31*(2), 161-175.

Pope, K. O., Sheffner, E. J., Linthicum, K. J., Bailey, C. L., Logan, T. M., Kasischke, E. S., ... & Roberts, C. R. (1992). Identification of central Kenyan Rift Valley fever virus vector habitats with Landsat TM and evaluation of their flooding status with airborne imaging radar. *Remote Sensing of Environment,40*(3), 185-196.

Reisen, W. K. (1995). Effect of temperature on Culex tarsalis (Diptera: Culicidae) from the Coachella and San Joaquin valleys of California. *Journal of Medical Entomology, 32*(5), 636-645.

Renner, I. W., & Warton, D. I. (2013). Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics,69*(1), 274-281.

Rogers, D. S., Belk, M. C., Gonzalez, M. W., Coleman, B. L., & Edwards, C. W. (January 01, 2006). Patterns of Habitat Use by Bats Along a Riparian Corridor in Northern Utah. *The Southwestern Naturalist, 51,* 1, 52-58.

Rios, J., Hacker, C., Hailey, C.A. & Parsons, R.E. (June 01, 2006). Demographic and Spatial Analysis Of West Nile Virus And St. Louis Encephalitis In Houston, Texas. *Journal of the American Mosquito Control Association, 22,* 2, 254-263.

Rogers, D. J. (2000). Satellites, space, time and the African trypanosomiases.*Advances in parasitology*, *47*, 129-171.

Rueda, L. M., Patel, K. J., Axtell, R. C., & Stinner, R. E. (1990). Temperature-dependent development and survival rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae). *Journal of medical entomology*, *27*(5), 892-898.

Shannon, C. E. (1948). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on information theory*, *26*, 26-37.

Stockwell, D. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science*, *13*(2), 143-158.

Stockwell, D. R., & Noble, I. R. (1992). Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, *33*(5), 385-390.

Ter Braak, C. J. (1987). The analysis of vegetation-environment relationships by canonical correspondence analysis. In *Theory and models in vegetation science*(pp. 69-77). Springer Netherlands.

Thomson, M. C., Connor, S. J., Milligan, P. J., & Flasse, S. P. (1996). The ecology of malaria--as seen from Earth-observation satellites. *Annals of tropical medicine and parasitology*, *90*(3), 243-264.

Thomson, M. C., Connor, S. J., Milligan, P., & Flasse, S. P. (1997). Mapping malaria risk in Africa: What can satellite data contribute?. *Parasitology Today*,*13*(8), 313-318.

Thuiller, W. (2003). BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*,*9*(10), 1353-1362.

Thuiller, W., Lavorel, S., & Araújo, M. B. (2005). Niche properties and geographical extent as predictors of species sensitivity to climate change.*Global Ecology and Biogeography*, *14*(4), 347-357.

Tran, A., Gardon, J., Weber, S., & Polidori, L. (2002). Mapping disease incidence in suburban areas using remotely sensed data. *American journal of epidemiology*, *156*(7), 662-668.

Turell, M. J., O'Guinn, M. L., Dohm, D. J., Webb, J. P., & Sardelis, M. R. (September 01, 2002). Vector Competence of Culex tarsalis from Orange County, California, for West Nile Virus. *Vector Borne and Zoonotic Diseases, 2,* 3, 193-196.

Vinogradova, E. B. (2000). *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetic, applied importance and control* (No. 2). Pensoft Publishers.

Wang, J., Ogden, N. H., & Zhu, H. (2011). The impact of weather conditions on Culex pipiens and Culex restuans (Diptera: Culicidae) abundance: A case study in Peel region. *Journal of medical entomology*, *48*(2), 468-475.

Ward, R. D. (January 01, 2005). Identification and Geographical Distribution of the Mosquitoes of North America, North of Mexico, 2nd Edition. By R. F. Darsie Jr. and R. A. Ward, pp. 416. University Press of Florida, USA, 2005. ISBN0 8130 2784 5. US$ 75.00. *Parasitology, 131,* 4.)

West Nile Virus." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 24 Sept. 2013. Web. 30 Sept. 2013. <http://www.cdc.gov/westnile/index.html>.

Wood, B., Washino, R., Beck, L., Hibbard, K., Pitcairn, M., Roberts, D., ... & Legters, L. (1991). Distinguishing high and low anopheline-producing rice fields using remote sensing and GIS technologies. *Preventive Veterinary Medicine*,*11*(3), 277-288.

Wood, B. L., Beck, L. R., Washino, R. K., Palchick, S. M., & Sebesta, P. D. (1991). Spectral and spatial characterization of rice field mosquito habitat. *International Journal of Remote Sensing*, *12*(3), 621-626.

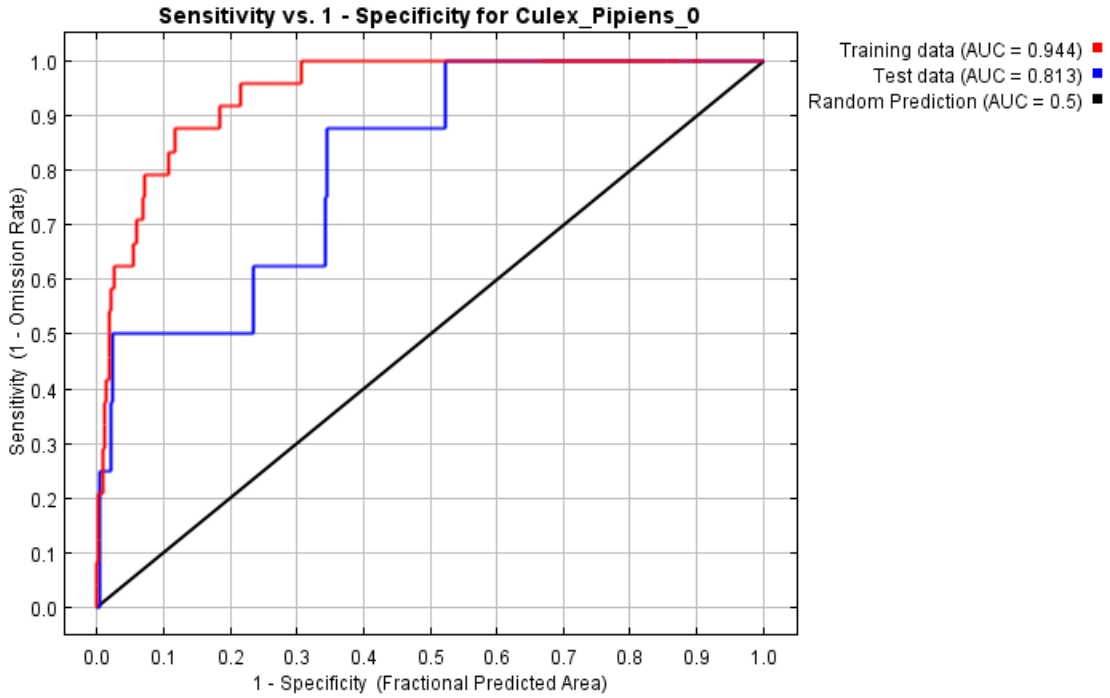
Wood, B. L., Beck, L. R., Washino, R. K., Hibbard, K. A., & Salute, J. S. (1992). Estimating high mosquito-producing rice fields using spectral and spatial data. *International Journal of Remote Sensing*, *13*(15), 2813-2826.

APPENDIX A

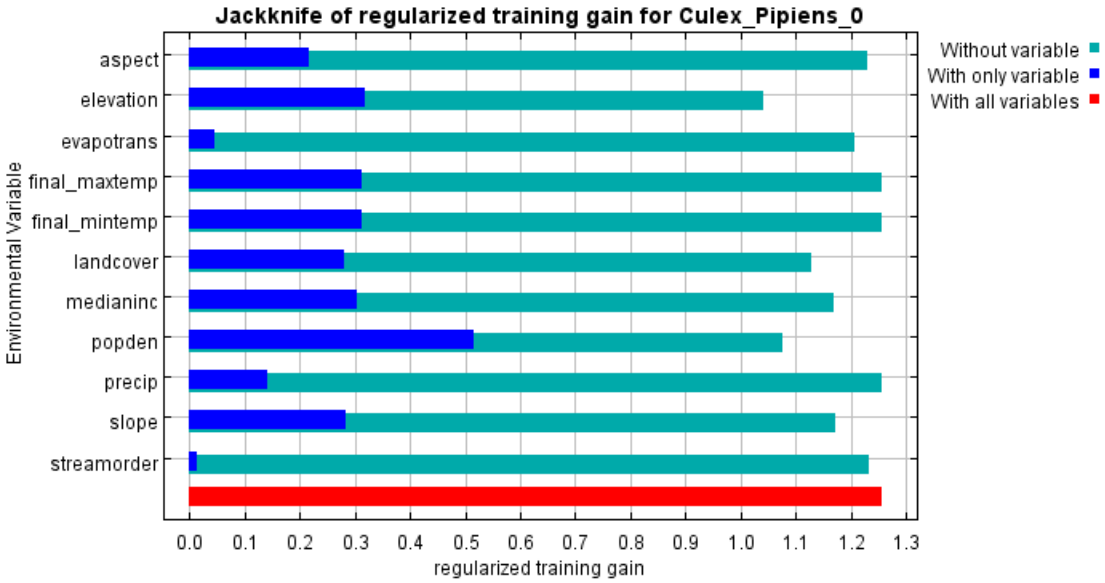
MAXENT MODELING REPLICANTS

A total of ten MaxEnt modeling replicates were produced in this study. The following are the outputs of each replicate.

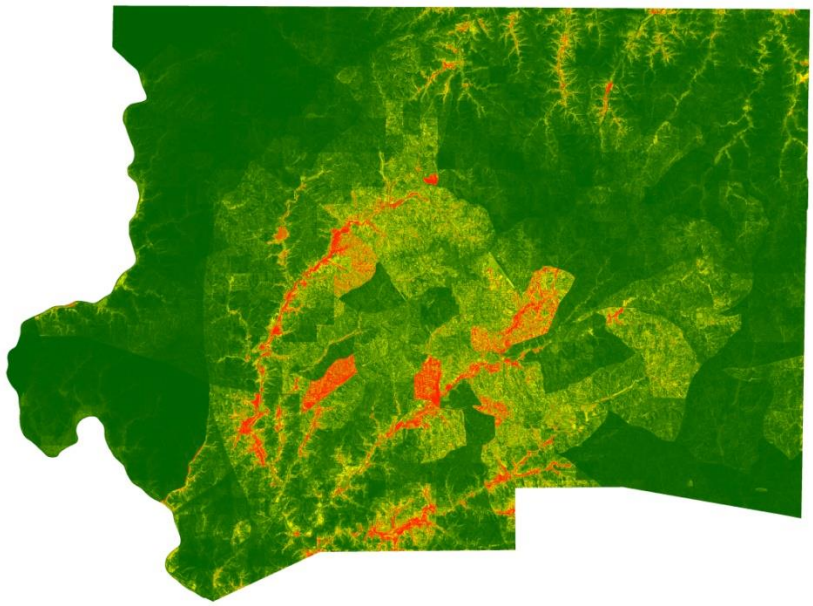
Replicate 1:



1a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

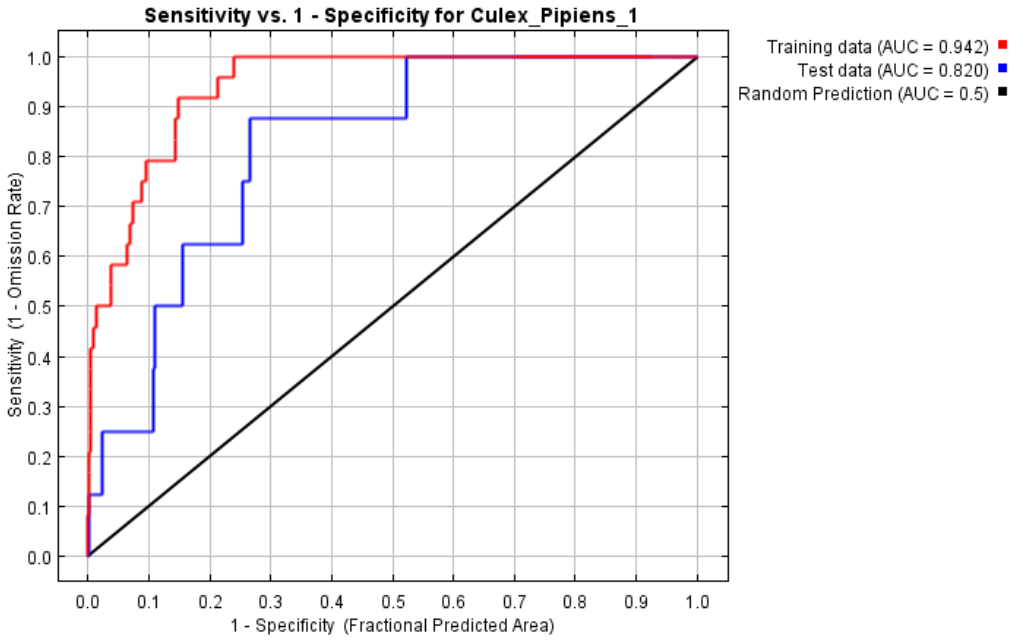


1b: Jackknife of regularized training gain for *Cx. pipiens*

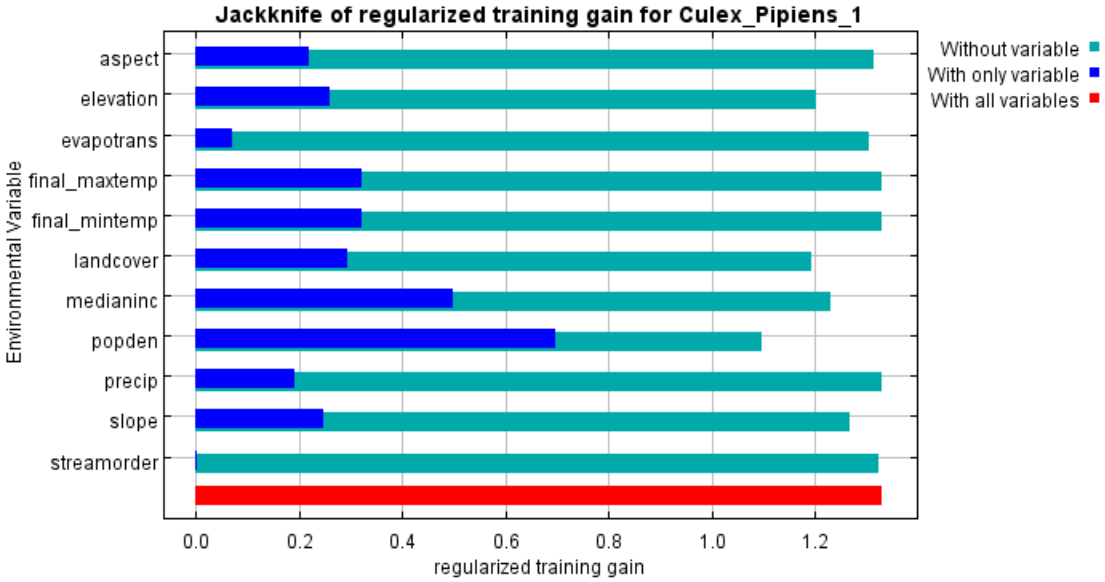


1c: Predictive Model for *Cx. pipiens*

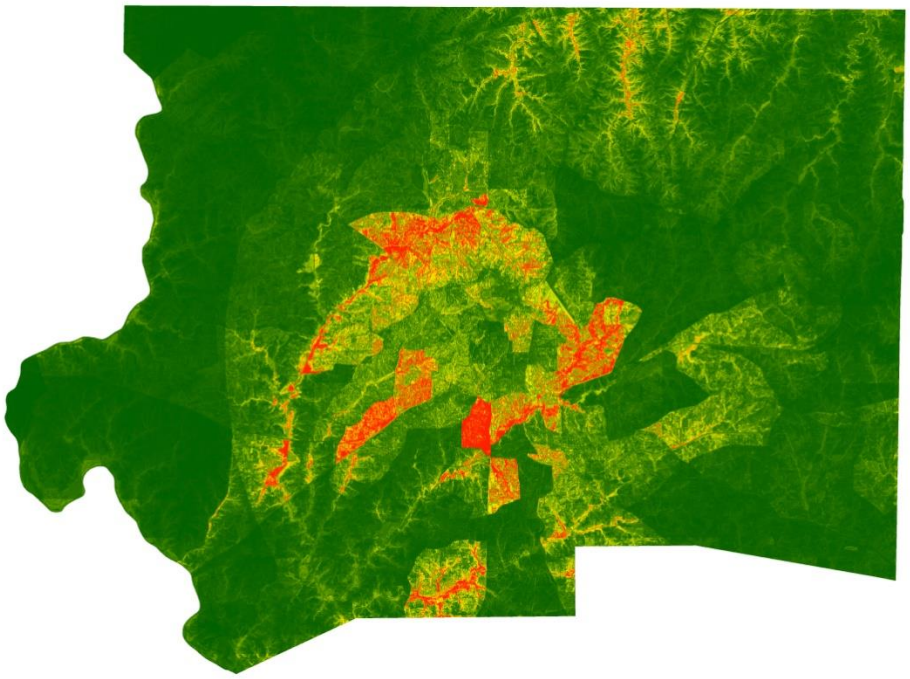
Replicate 2:



2a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

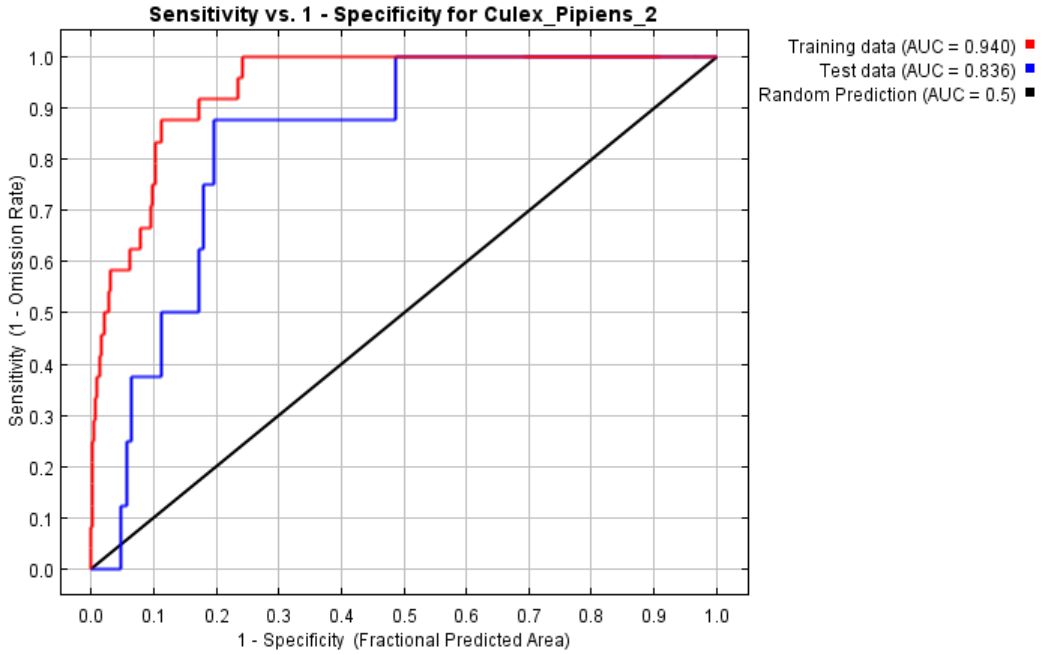


2b: Jackknife of regularized training gain for *Cx. pipiens*

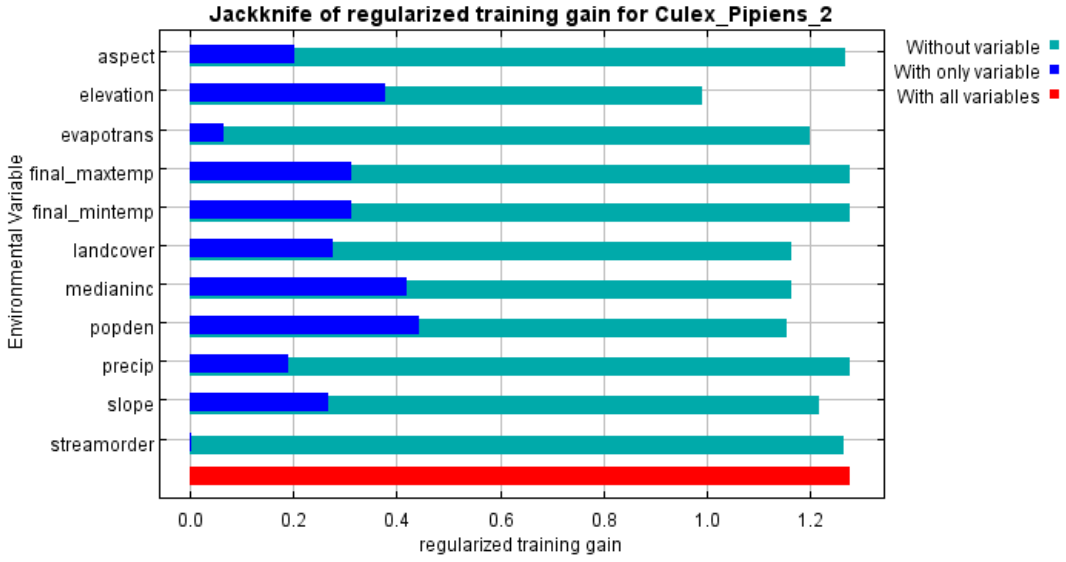


2c: Predictive Model for *Cx. pipiens*

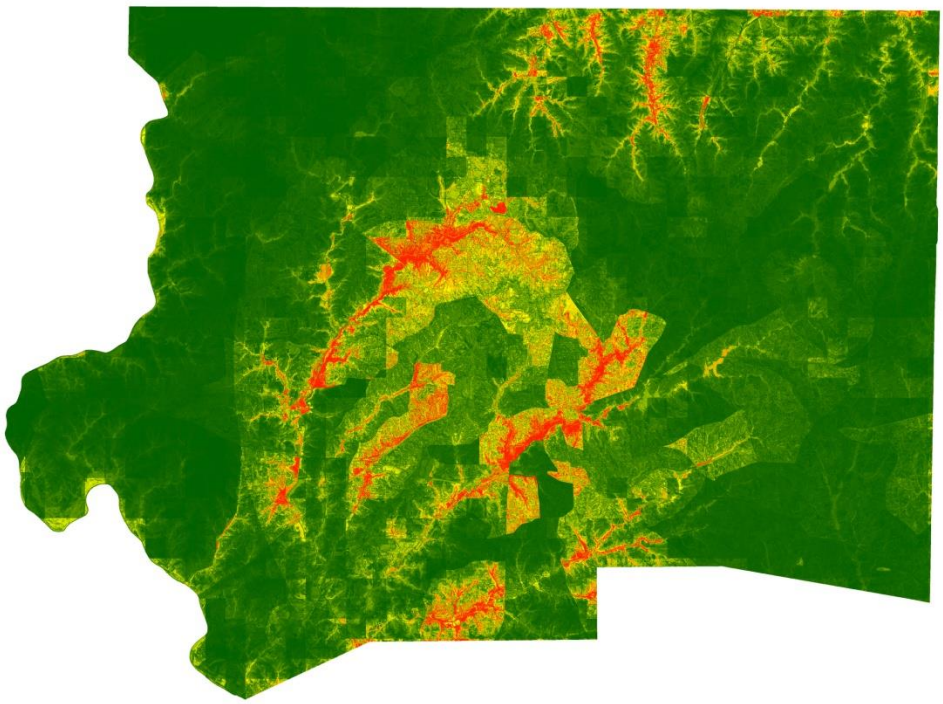
Replicate 3:



3a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

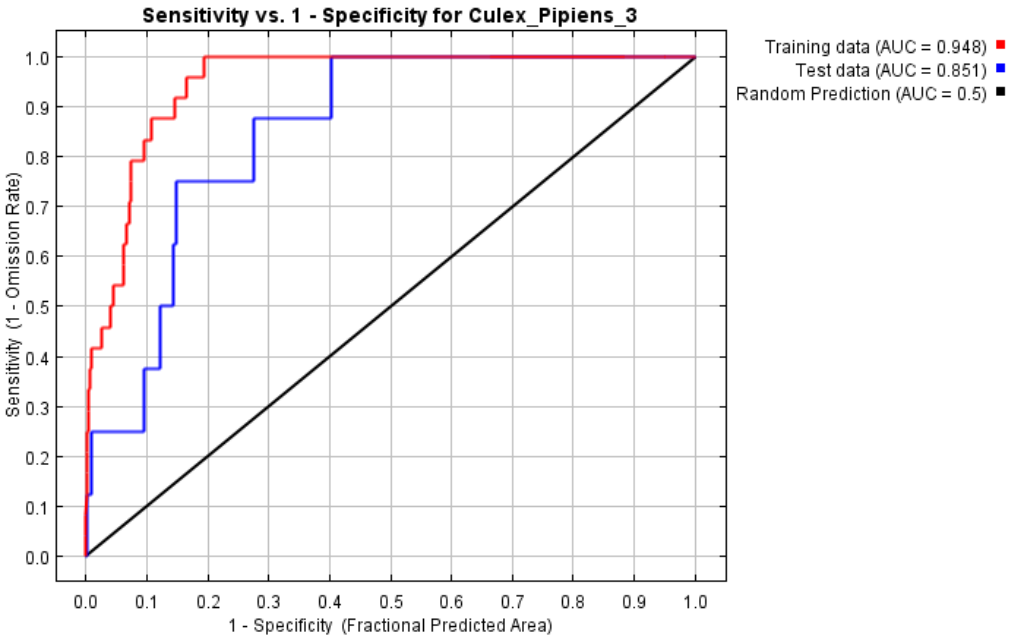


3b: Jackknife of regularized training gain for *Cx. pipiens*

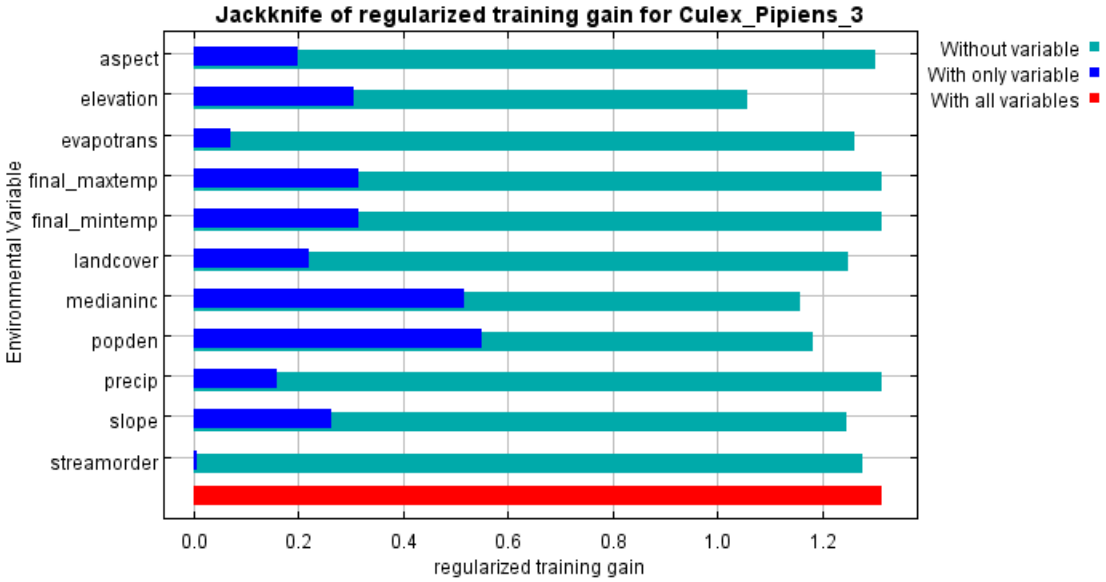


3c: Predictive Model for *Cx. pipiens*

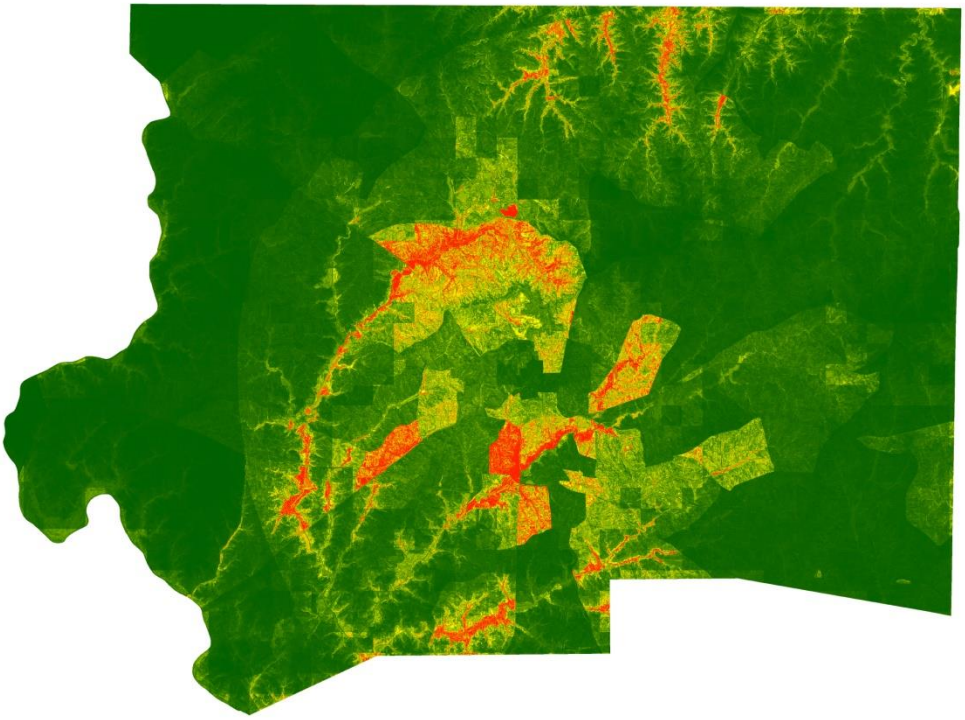
Replicate 4:



4a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

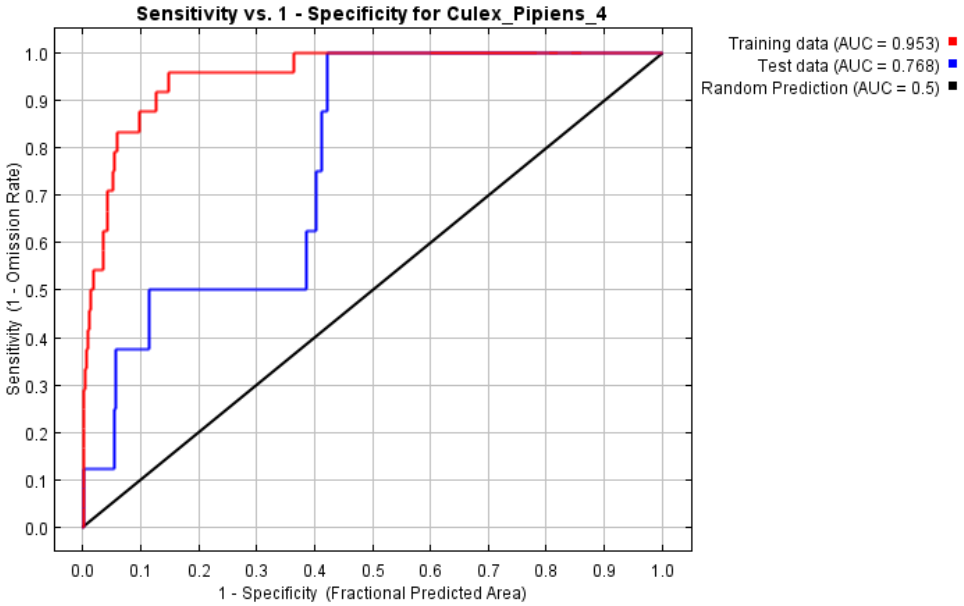


4b: Jackknife of regularized training gain for *Cx. pipiens*

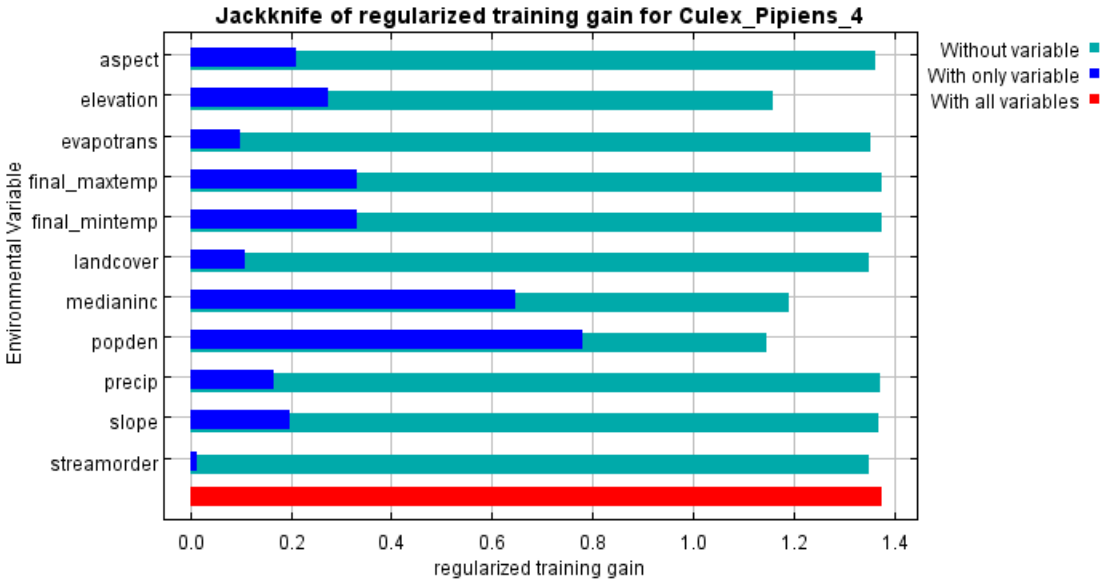


4c: Predictive Model for *Cx. pipiens*

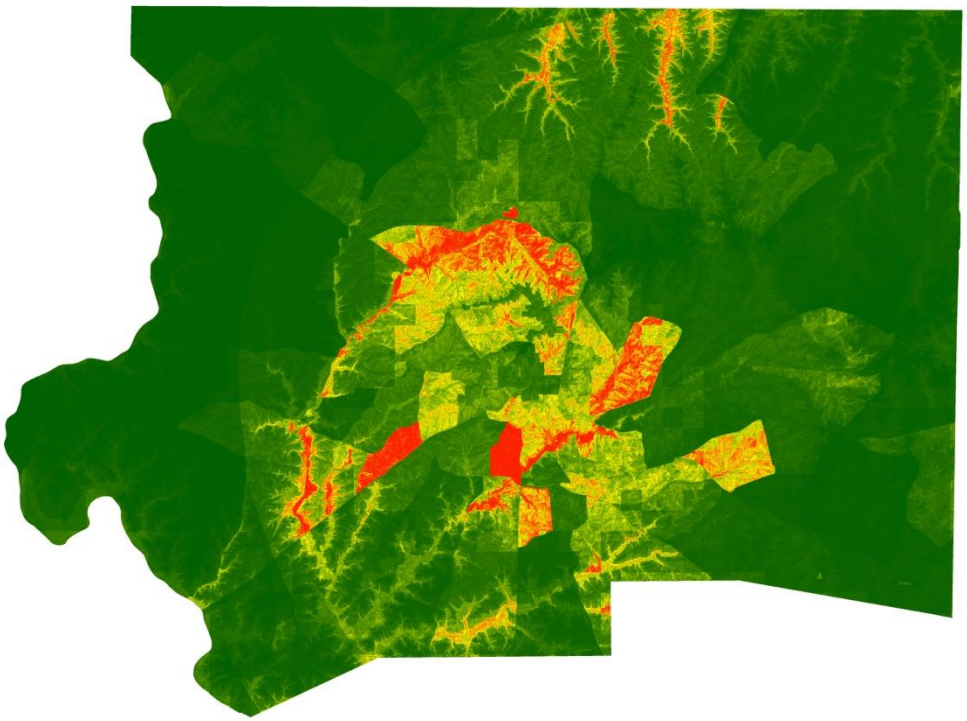
Replicate 5:



5a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

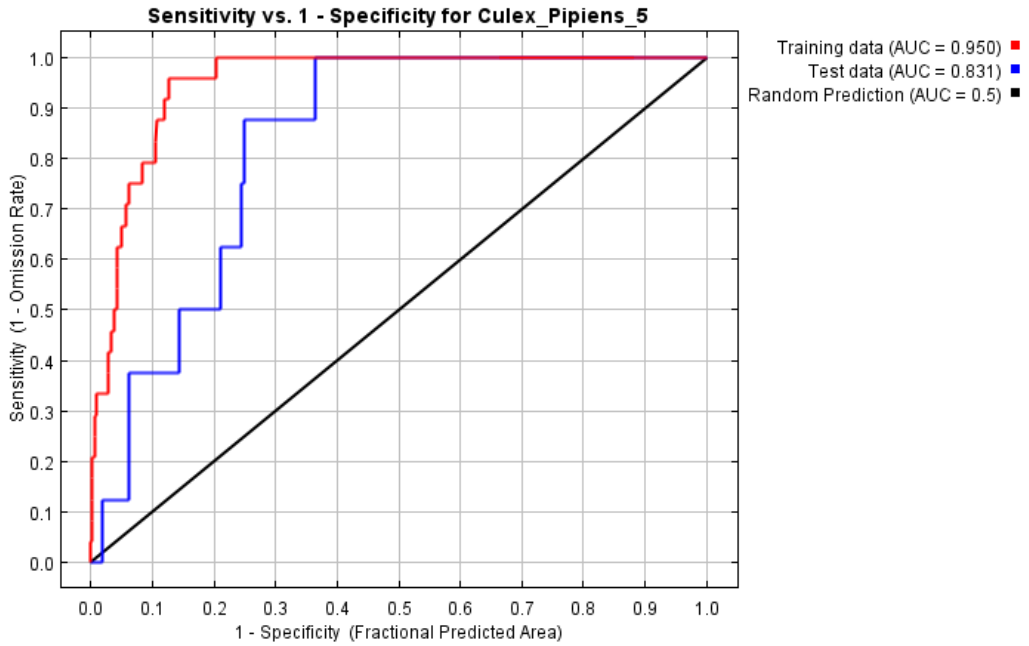


5b: Jackknife of regularized training gain for *Cx. pipiens*

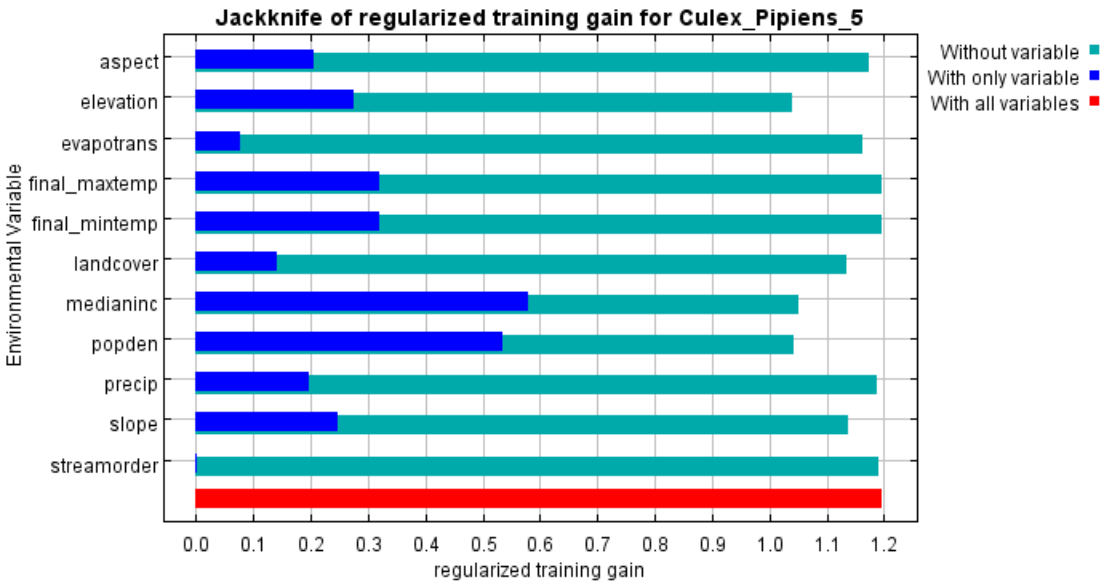


5c: Predictive Model for *Cx. pipiens*

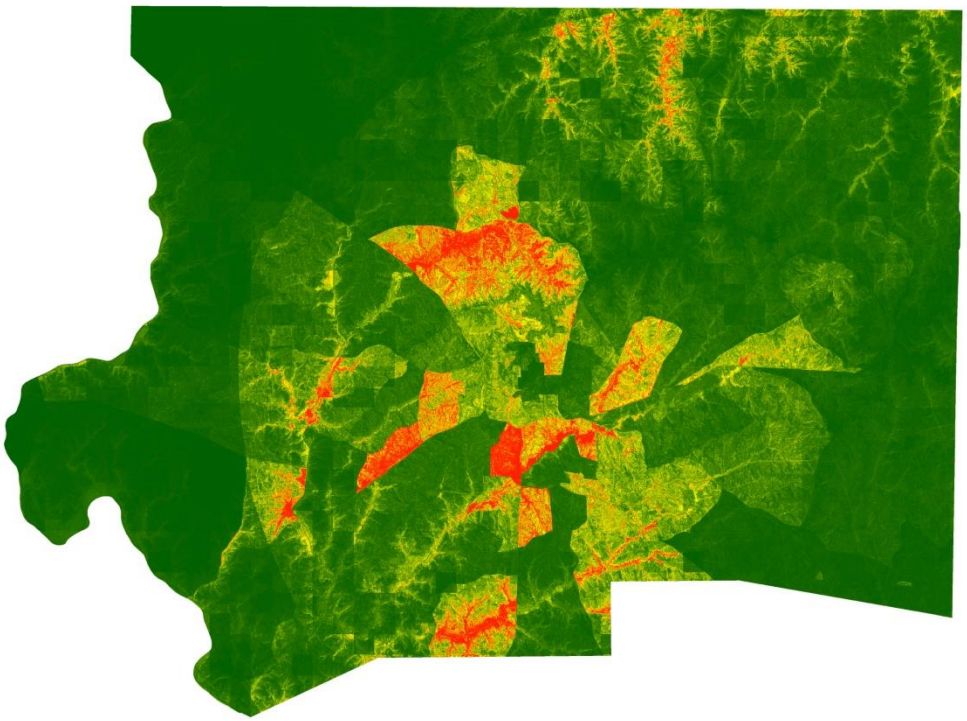
Replicate 6:



6a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

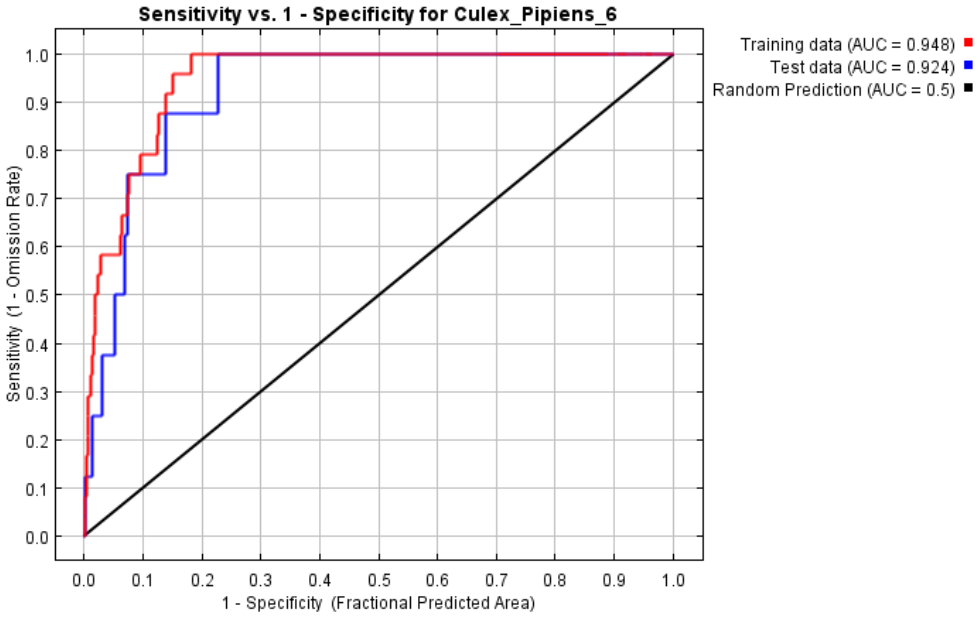


6b: Jackknife of regularized training gain for *Cx. pipiens*

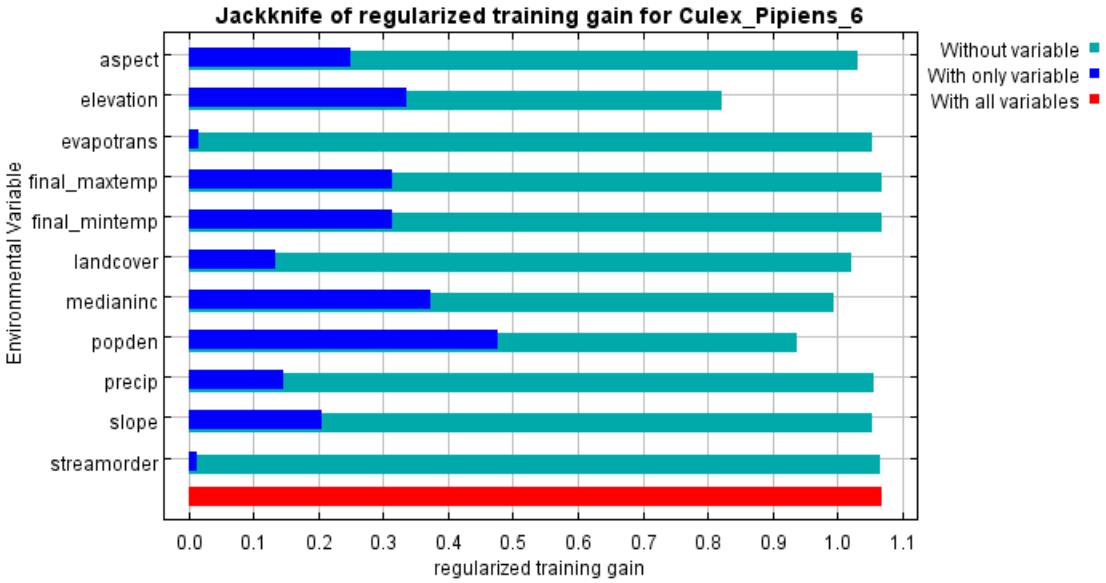


6c: Predictive Model for *Cx. pipiens*

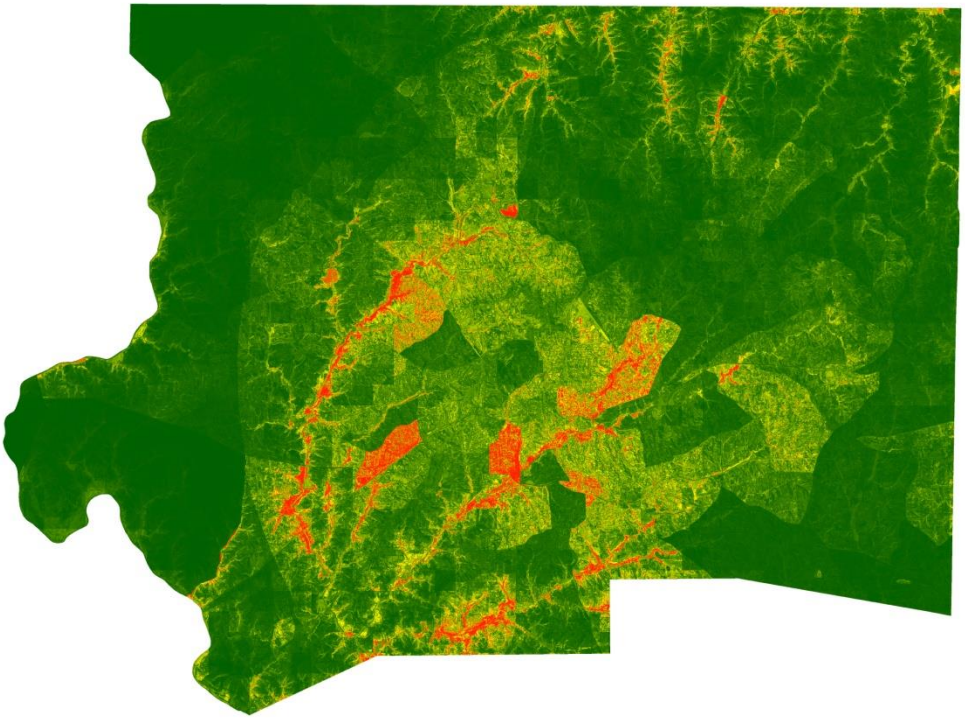
Replicate 7:



7a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

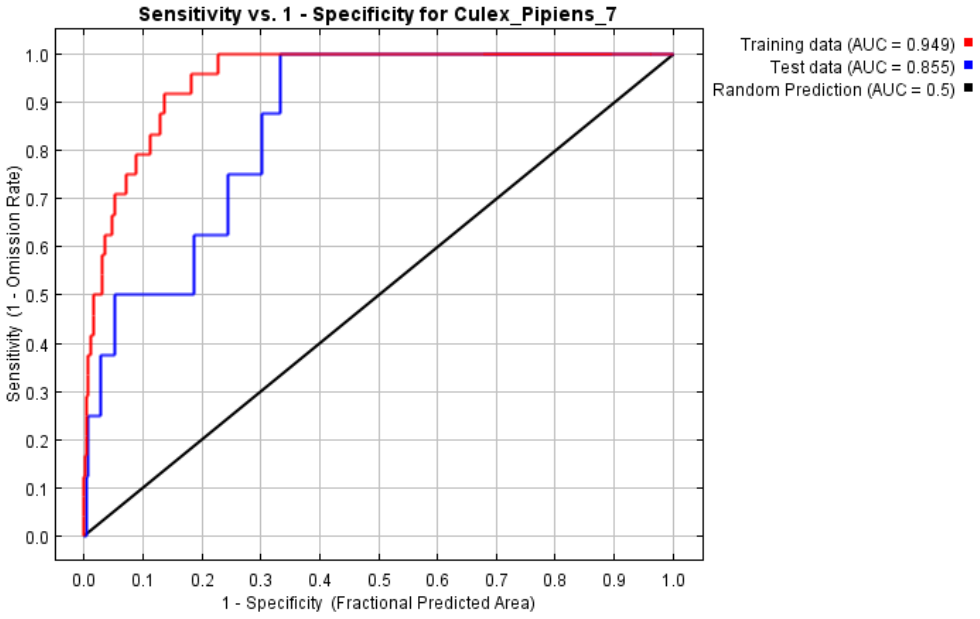


7b: Jackknife of regularized training gain for *Cx. pipiens*

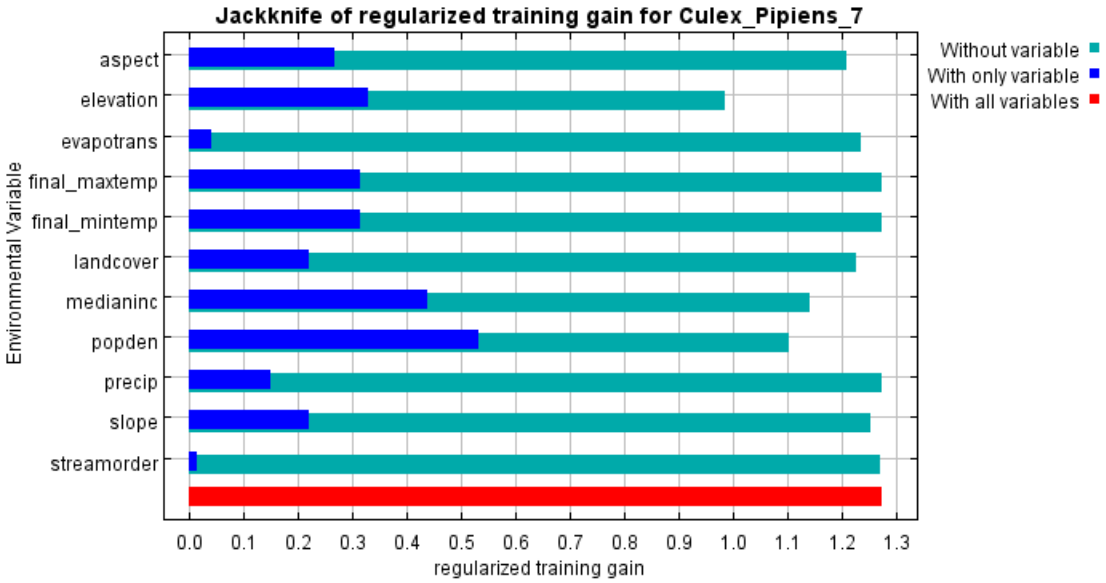


7c: Predictive Model for *Cx. pipiens*

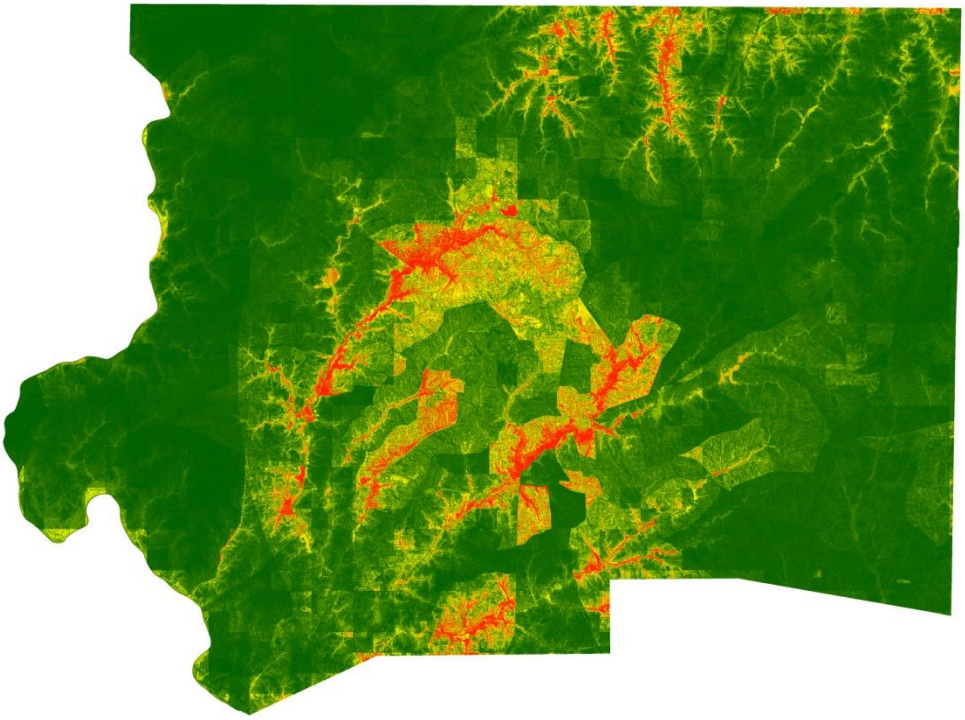
Replicate 8:



8a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

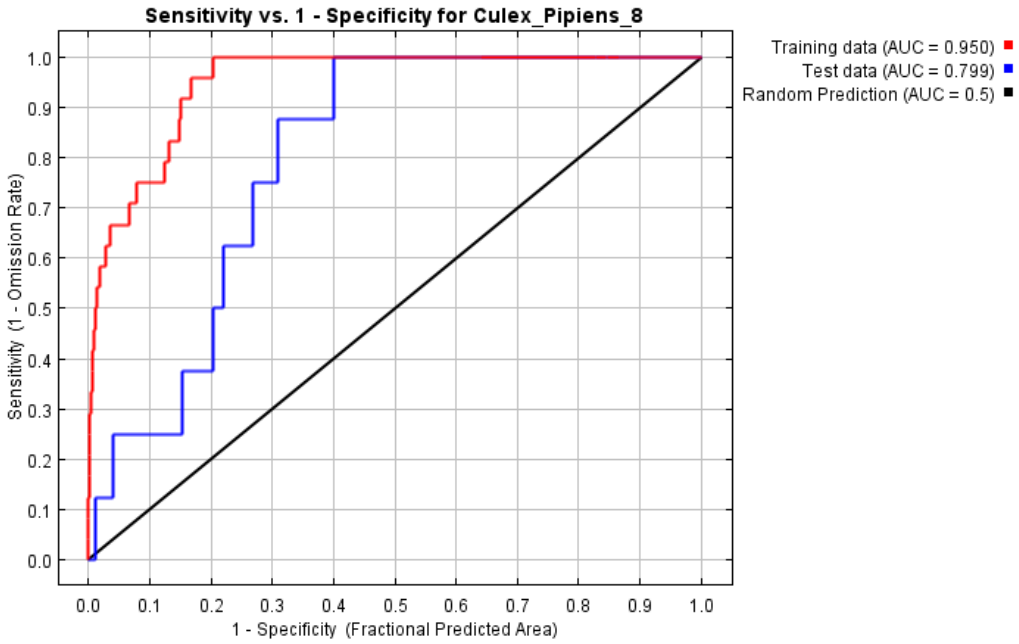


8b: Jackknife of regularized training gain for *Cx. pipiens*

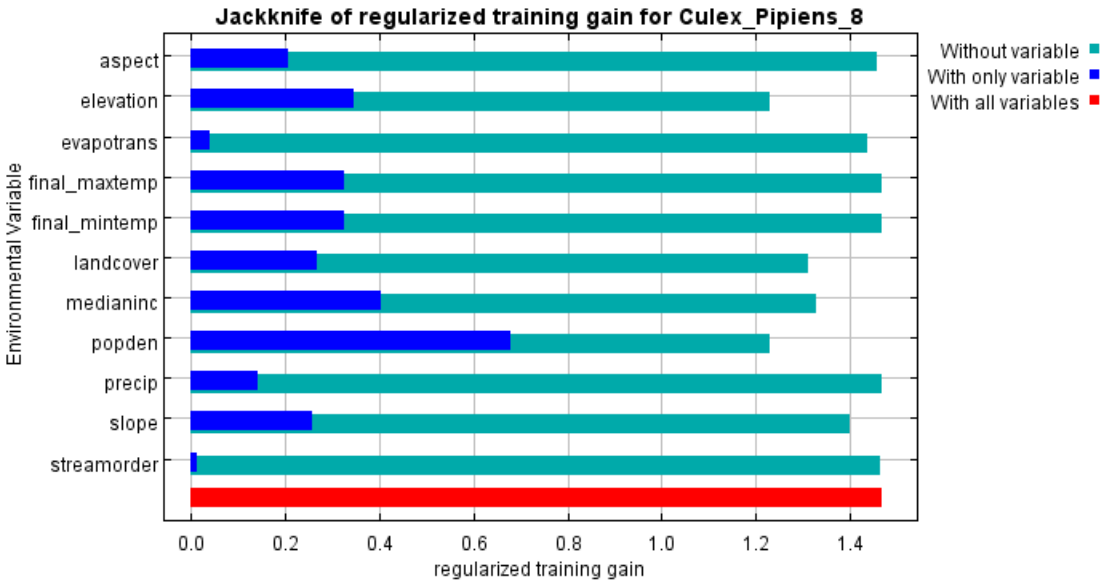


8c: Predictive Model for *Cx. pipiens*

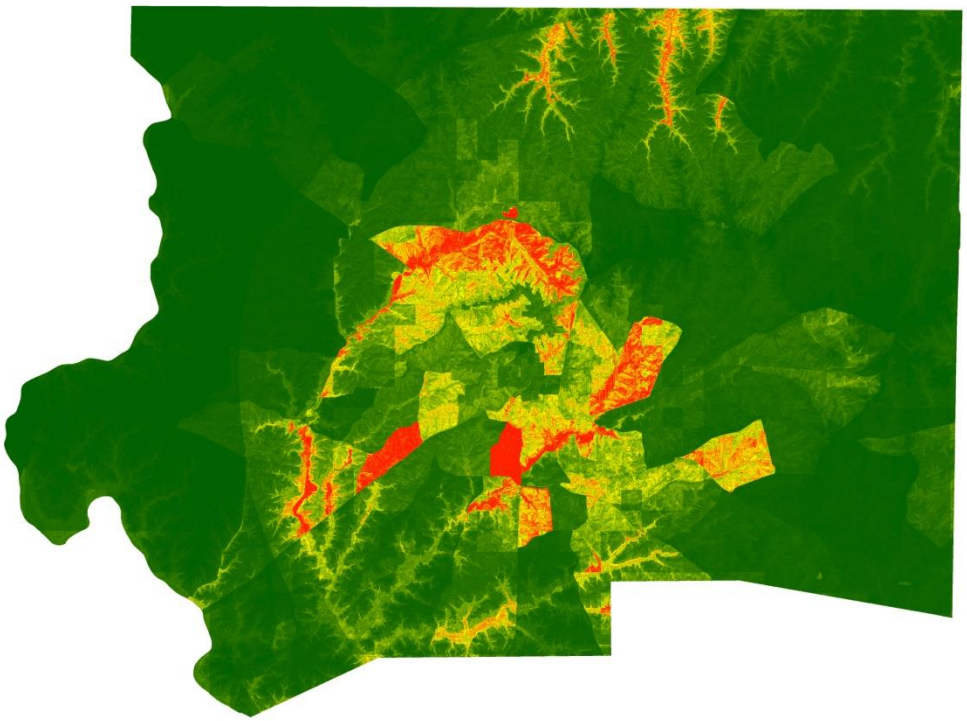
Replicate 9:



9a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.

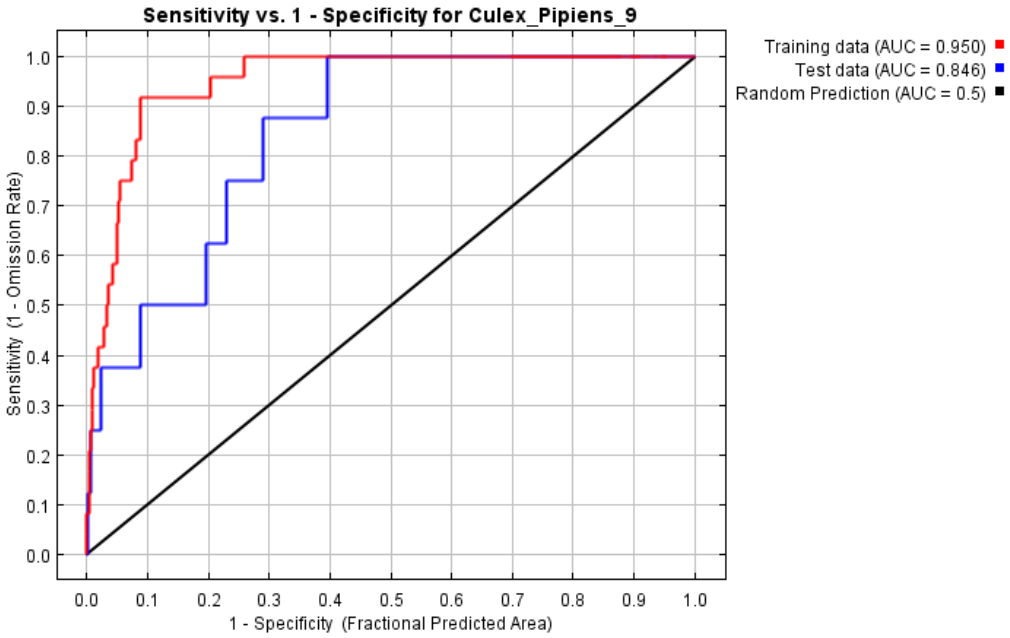


9b: Jackknife of regularized training gain for *Cx. pipiens*

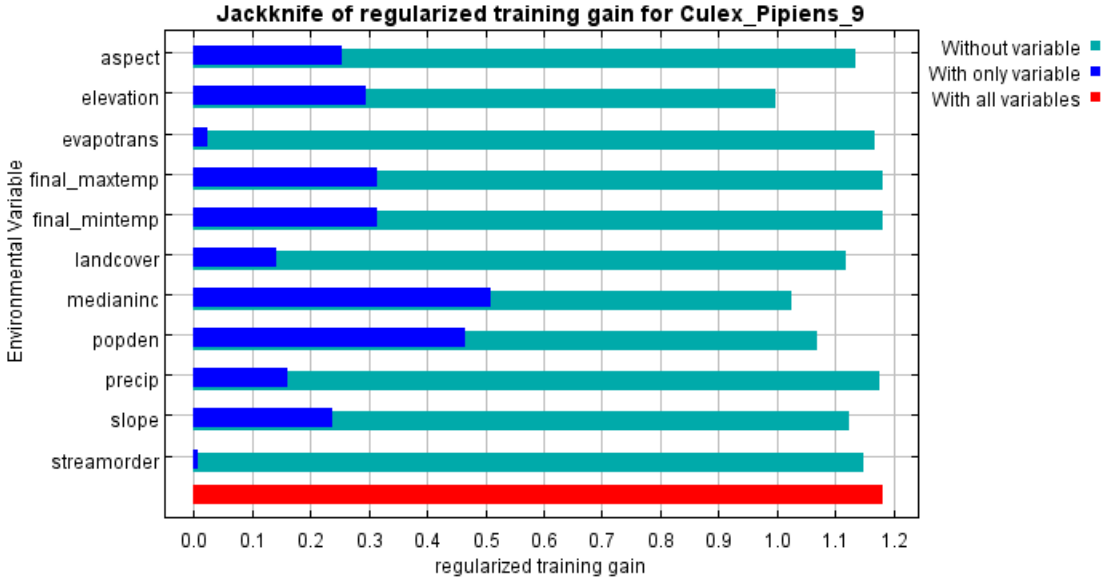


9c: Predictive Model for *Cx. pipiens*

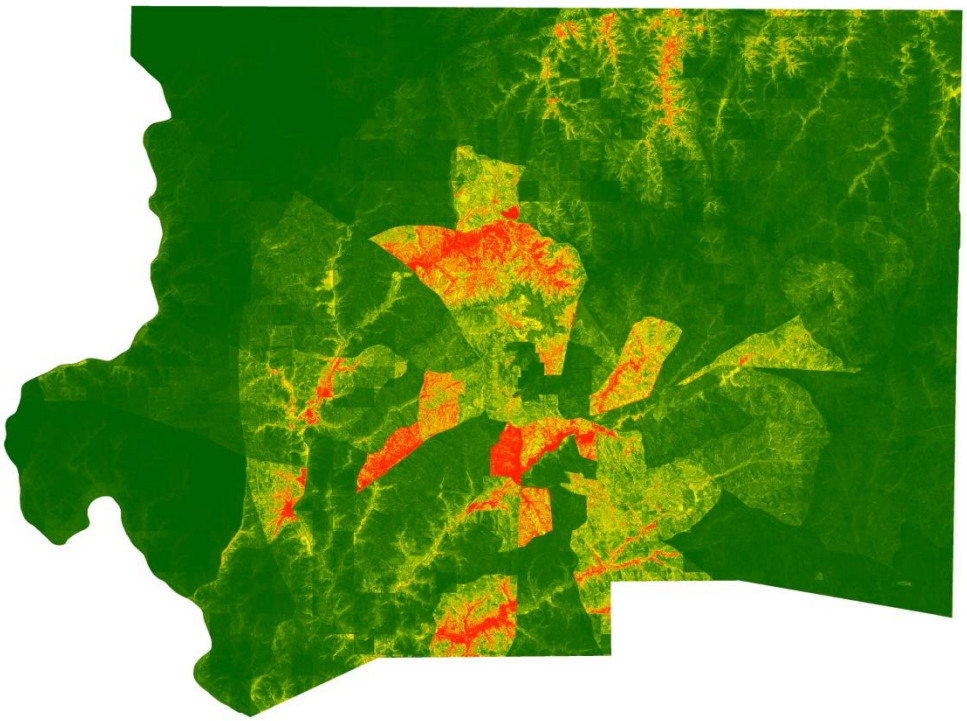
Replicate 10:



10a: Receiver Operating Curve (ROC) for both training and test data for *Cx. Pipiens*.



10b: Jackknife of regularized training gain for *Cx. pipiens*



10c: Predictive Model for *Cx. pipiens*